

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:12 ; Search time 71.8824 Seconds  
(without alignments)  
4014.538 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_109000\_109519

Perfect score: 520  
Sequence: 1 ggtgcagcagcagcagcgg.....ggccagcgccgcgggatcc 520

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : Issued Patents\_NA.\*

1: /cgm2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgm2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgm2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgm2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgm2\_6/ptodata/2/ina/PCTUS.COMB.seq.\*  
6: /cgm2\_6/ptodata/2/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	55.2	10.6	4411529	3	US-09-103-840A-1
C 2	49.2	9.5	2061	4	US-09-252-991A-3938
C 3	49.2	9.5	2316	4	US-09-252-991A-3884
C 4	49.2	9.5	2328	4	US-09-252-991A-3965
C 5	49	9.4	765	4	US-09-252-991A-3980
C 6	49	9.4	2223	4	US-09-252-991A-4015
C 7	48.6	9.3	816	4	US-09-252-991A-13914
C 8	48.6	9.3	966	4	US-09-252-991A-13437
C 9	48.6	9.3	1668	4	US-09-252-991A-13705
C 10	48.6	9.3	4403765	3	US-09-103-840A-2
C 11	48.6	9.3	4403765	3	US-09-103-840A-2
C 12	48	9.2	606	4	US-09-252-991A-3898
C 13	47	9.0	1835	4	US-09-417-704-2
C 14	47	9.0	4411529	3	US-09-103-840A-1
C 15	46.8	9.0	1977	4	US-09-833-381-1616
C 16	45.4	8.7	1461	4	US-09-252-991A-180
C 17	45.4	8.7	1560	4	US-09-252-991A-166
C 18	45.4	8.7	2610	4	US-09-252-991A-175
C 19	45.2	8.7	23673	4	US-09-773-816-1
C 20	44.8	8.6	1125	4	US-09-252-991A-6994
C 21	44.8	8.6	1929	4	US-09-252-991A-7017
C 22	44.8	8.6	1947	4	US-09-252-991A-7111
C 23	44.6	8.6	1614	4	US-09-616-289-45
C 24	44.6	8.6	1839	4	US-09-252-991A-1273
C 25	44.6	8.6	1980	4	US-09-252-991A-1333
C 26	44.6	8.6	2561	4	US-09-616-289-48
C 27	44.6	8.6	12425	4	US-09-616-289-50

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37Rv

US-09-103-840A-1

Query Match 10.6%; Score 55.2; DB 3; Length 4411529;

Best Local Similarity 47.5%; Pred. No. 0.0088; Indels 1; Gaps 1;  
Matches 194; Conservative 0; Mismatches 213;

QY	109	CCTGCCGCGCATGCCGAAACACCGGGCCAGGATCACGTGCAGCAGCGCGATGCGCGCT 168
DB	338830	CCTCCGCTGGCGCGGTACG 338771
QY	169	CGATCTCGAGGTCG 228
DB	338770	CGATCAACGAGTTCTTCTTCTGGCGAATACCGGGCGCGCGCGCGCGCGCGCGCGCGCG 338711
QY	229	GGCTCGGTGCGACGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288
DB	338710	GCGCCCG 338651
QY	289	CTCGCGGTCGCGAGCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 348
DB	338650	CCGAGAGATCCG 338591
QY	349	TGCGGTCGTCGCCCTTGGCTTGGAGGATAGCGGTTTACGACGAGCGCGCGCGCGCGCG 408





;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
;; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
;; FILE REFERENCE: 107196.136  
;; CURRENT APPLICATION NUMBER: US/09/252,991A  
;; PRIOR FILING DATE: 1999-02-18  
;; PRIOR FILING DATE: 1999-02-18  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR FILING DATE: 1998-02-18  
;; PRIOR FILING DATE: 1998-07-27  
;; NUMBER OF SEQ ID NOS: 33142  
;; SEQ ID NO 3980  
;; LENGTH: 765  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3980

Query Match 9.4%; Score 49; DB 4; Length 765;  
Best Local Similarity 45.2%; Pred. No. 0.14;  
Matches 219; Conservative 0; Mismatches 265; Indels 1; Gaps 1;  
QY 1 GGTGACAGCAACGACCGGTGGTGGCGCGCGCGCCATCAGCAACTGCCACTCCGGCG 60  
DB 64 GGGTCACTACCATCAGGTTGGCGCGCGCGCAACCGGAGAGGTTGCGCAGCACCTGG 123  
QY 61 GGGTACAGGTGACAGAGTGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGCGCA 120  
DB 124 CCTCGCTGGCGCGCTCAGAGGCTGGCGGTGGGCTTTGTCGAGCAGCAGCAACGGGCGCGTCG 183  
QY 121 TGCCGAAACACCGCGCCAGGATCACGTGACAGCGGCGATGCGGCGCTCGCATCTCGACGG 180  
DB 184 AGACAGCCCGGAGCCAGGACAGGCGTTGCGCGTGGCGCGGAGCAGCAGCGCGCGC 243  
QY 181 TCGCGCGAGTGCATCTGTCGCGCGCTCCACACAGGGGAAATGCGCTCGGTGGCA 240  
DB 244 CGGACCGGACCTGCTGCCAGCGCTCGCGCCAGGCTCGATCTCTTCCAGCAGGCCAA 303  
QY 241 GCGCAGCCCGCAGCGGAGAGCTGTCAGGCTGGGCTGACGAGCTCTCGCGGTCGG 300  
DB 304 CGGCGCGCGCGCTAGCGAGCGCTTCGAGATCGGCGTTCGCTCGCGCCTCGCAGGT 363  
QY 301 GAGCGGAAACGCGCAGCGCTCAGATCCCTGTGCTAGTCGATCGGCTCAGTCCGCTGCTCC 360  
DB 364 TCCAGGCAACCTGCGCGGAAAGGCGCTGTCTCGAAC-ACCAGGTTACGATTGGCG 422  
QY 361 CCCTTGGCTGGGAGATAGCGTTACAGCAGCGGACACACGCGCGCGGCGGGG 420  
DB 423 CGAGGGTCTGTTTCGCTCAACCGCGGATATCCACGCCACCCAGCAGGACACGCGCGCA 482  
QY 421 CGGCTTACGCGATCCGCTCGATGACGAGCGGCTGCGGGGTCGGGCGGCTCGCGGATC 480  
DB 483 TCGAGCTGTAGAGCGCGCGGAGCAGCGCCAGCAGGCTGCTCTTGGCGCGCGGAGGT 542  
QY 481 CGTAC 485  
DB 543 CCGAC 547

RESULT 6  
US-09-252-991A-4015  
; Sequence 4015, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

;; SEQ ID NO 4015  
;; LENGTH: 2223  
;; TYPE: DNA  
;; ORGANISM: Pseudomonas aeruginosa  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (2110)  
;; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-4015

Query Match 9.4%; Score 49; DB 4; Length 2223;  
Best Local Similarity 45.2%; Pred. No. 0.14;  
Matches 219; Conservative 0; Mismatches 265; Indels 1; Gaps 1;  
QY 1 GGTGACAGCAACGACCGGTGGTGGCGCGCGCCATCAGCAACTGCCACTCCGGCG 60  
DB 615 GGGTCACTACCATCAGGTTGGCGCGCGCGCAACCGGAGAGGTTGCGCAGCACCTCG 674  
QY 61 GGGTACAGGTGACAGAGTGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGCGCA 120  
DB 675 CCTCGCTGGCGCGCTCAGAGGCTGGCGGTGGGCTTTGTCGAGCAGCAGCAACGGGCGCGTCG 734  
QY 121 TGCCGAAACACCGCGCCAGGATCACGTGACAGCGGCGATGCGGCGCTCGATCTCGACGG 180  
DB 735 AGACAGCCCGGAGCCAGGACAGGCGTTGCGCGTGGCGCGGAGCAGCAGCGCGCGC 794  
QY 181 TCGCGCGAGTGCATCTGTCGCGCGCTCCACACAGGGGAAATGCGCTCGGTGGCA 240  
DB 795 CGGACCGGACCTGCTGCCAGCGCTCGCGCCAGGCTCGATCTCTTCCAGCAGGCCAA 854  
QY 241 GCGCAGCCCGCAGCGGAGAGCTGTCAGGCTGGGCTGACGAGCTCTCGCGGTCGG 300  
DB 855 CGGCGCGCGCGCTAGCGCAGCGCTTCGAGATCGGCGTTCGCTCGCGCCTCGCAGGT 914  
QY 301 GAGCGGAAACGCGCAGCGCTCAGATCCCTGTGCTAGTCGATCGGCTCAGTCCGCTCGTCC 360  
DB 915 TCCAGGCAACGCTGCGCGGAGAGCGCGCTGTCTCGTCCGAG-ACCAGGTTACGATTGGCG 973  
QY 361 CCCTTGGCTGGGAGATAGCGGTTACAGCAGCGGACACACGCGCGCGGCGCGGG 420  
DB 974 GCGAGGGTCTGTTTCGCTCAACCGCGGATATCCACGCCACCCAGCAGGACACGCGCGCA 1033  
QY 421 GCGGTTACGCGATCCGCTCGATGACAGCGGCTCGGGGTCGGGCGGTCGGCGGATC 480  
DB 1034 TCGAGCTGTAGAGCGCGCGGAGCAGCGCCAGCAGGCTGCTCTTCCGCGCGCGGAGGT 1093  
QY 481 CGTAC 485  
DB 1094 CCGAC 1098

RESULT 7  
US-09-252-991A-13914/c  
; Sequence 13914, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13914  
; LENGTH: 816  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13914

Query Match 9.3%; Score 48.6; DB 4; Length 816;  
Best Local Similarity 52.6%; Pred. No. 0.16;  
Matches 133; Conservative 0; Mismatches 114; Indels 6; Gaps 1;

QY 65 GACCAGGTTCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCTGCGCGGATGCC 124  
DB 661 GACCAGGTTCAGTTCGTGGTGTATCAGCAGCAGGACATGCCAGCGTGCCTGAGGTC 602

QY 125 GAAACACCGGCGCAGGATACGTCGACAGCAGCGGATGCGCGCTCGATCTGACAGGTCGG 184  
DB 601 CTTGAGCAGATCAGGATCTTCACTGACGCTG-----CGTCAGCGCGTGTGTCGG 548

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACACAGCGGAACTGCTCGTGGCAGCGG 244  
DB 547 TTGCTGCGCGATCAGCAGTTGCGGCTCGCAGCGCGGATGCGCATCATCACCGCTG 488

QY 245 CAGCCCGCAGCGGACAGCTGCTCCAGGCTGCGGTCAAGGACTCTTCGCGGTCGCGAGC 304  
DB 487 CCGCTGCGCGCGCAGAGTTGCTGTGATAGGCTTTTCAGGCGCTTGGCGCGCTCGGGAT 428

QY 305 GGAACGCGCACG 317  
DB 427 GCCGACAGCTCG 415

RESULT 8  
US-09-252-991A-13437  
; Sequence 13437, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13437  
; LENGTH: 966  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13437

Query Match 9.3%; Score 48.6; DB 4; Length 966;  
Best Local Similarity 52.6%; Pred. No. 0.16;  
Matches 133; Conservative 0; Mismatches 114; Indels 6; Gaps 1;

QY 65 GACCAGGTTCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCTGCGCGGATGCC 124  
DB 262 GACCAGGTTCAGTTCGTGGTGTATCAGCAGCAGGACATGCCAGCGTGCCTGAGGTC 321

QY 125 GAAACACCGGCGCAGGATACGTCGACAGCGGATGCGCGCTCGATCTGACAGGTCGG 184  
DB 322 CTTGAGCAGATCAGGATCTTCACTGACGCTG-----CGTCAGCGCGTGTGTCGG 375

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACACAGCGGAACTGCTCGTGGCAGCGG 244  
DB 376 TTGCTGCGCGATCAGCAGTTGCGGCTCGCAGCGCGGATGCGCATCATCACCGCTG 435

QY 245 CAGCCCGCAGCGGACAGCTGCTCCAGGCTGGGCTCAGGACTCTTCGCGGTCGCGAGC 304  
DB 436 CCGCTGCGCGCGCAGAGTTGCTGTGATAGGCTTTTCAGGCGCTTGGCGCGCTCGGGAT 495

QY 305 GGAACGCGCACG 317  
DB 496 GCCGACAGCTCG 508

RESULT 9

US-09-252-991A-13705/c  
; Sequence 13705, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13705  
; LENGTH: 1668  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13705

Query Match 9.3%; Score 48.6; DB 4; Length 1668;  
Best Local Similarity 52.6%; Pred. No. 0.16;  
Matches 133; Conservative 0; Mismatches 114; Indels 6; Gaps 1;

QY 65 GACCAGGTTCACCAAGTGGGAGTTGAGCAGCAGCTCATCGCGCTGCGCGGATGCC 124  
DB 717 GACCAGGTTCAGTTCGTGGTGTATCAGCAGCAGGACATGCCAGCGCTGCCTGAGGTC 658

QY 125 GAAACACCGGCGCAGGATACGTCGACAGCGGATGCGCGCTCGATCTGACAGGTCGG 184  
DB 657 CTTGAGCAGATCAGGATCTTCACTGACGCTG-----CGTCAGCGCGTGTGTCGG 604

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACACAGCGGAACTGCTCGTGGCAGCGG 244  
DB 603 TTGCTGCGCGATCAGCAGTTGCGGCTCGCAGCGCGGATGCGCATCATCACCGCTG 544

QY 245 CAGCCCGCAGCGGACAGCTGCTCCAGGCTGGGCTCAGGACTCTTCGCGGTCGCGAGC 304  
DB 543 CCGCTGCGCGCGCAGAGTTGCTGTGATAGGCTTTTCAGGCGCTTGGCGCGCTCGGGAT 484

QY 305 GGAACGCGCACG 317  
DB 483 GCCGACAGCTCG 471

RESULT 10  
US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 9.3%; Score 48.6; DB 3; Length 4403765;

Best Local Similarity 45.5%; Pred. No. 0.17;  
Matches 209; Conservative 0; Mismatches 249; Indels 1; Gaps 1;  
QY 57 GCGGGGTGACAGAGTCAGACAGGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGGC 116  
Db 3925089 GCGGTGATCAATGCGCCACACAGCGCTGCTGGGGCCCGTTGATCGGTGACGCGGC 3925148  
QY 117 GGCATCCGAAACACCGGCGCAGATCAGTGCAGCAGCGGATGCGGCGCTCGATTCG 176  
Db 3925149 AATGCGAGACTCCGCGGCGGGCCCGCGGCGCTGCTGCTGTTGCGCAACGCGGG 3925208  
QY 177 ACGGTCGCGCCAGCTCGATCTCGTCGCGCCCGCTCCACACAGCGGGAACCTGCTCGGT 236  
Db 3925209 CCGGGGCGAGCGGGGCGCCCGGCGAGCGCGGCGCTGCGGGCCCGCGGATTGCG 3925268  
QY 237 GCGAGCGCAGCCCGCAGCGGAGCAGTCTGTCAGGCTGGGCTCAGGACCTTCGCGGG 296  
Db 3925269 GGCACCGCGCGCGCGCGGCGCGCGGCGAGCGGTGGGGGCGACCGCGGTGCGCGGCG 3925328  
QY 297 TCGGAGCGGAAACCGCGCAGCGCTCAGATCCCTGTTCAGTCGATCGGCTCAGTCCGCGTC 356  
Db 3925329 CCGGTGGGTGGCTGTTGCGGTTGGCGGCGCCCGCGGTGTTCAGTGGGCGCGTGGCGGC 3925388  
QY 357 GTCCCGCTTGGCTGGGAGGATAGCGGTTCAGACAGCGGCGCACACAGCGGCGGGCGGC 416  
Db 3925389 ACCGCGCGGCGCGCGCGCGCGGTGTTGATCTCGGCGCGCGCGCGCGCGGTGTC 3925448  
QY 417 GGGGCGGTTCAGCGGATCGCTCGATGACAGCGGCTGCGGGTTCGGGCGGTTCGCGCG 476  
Db 3925449 GTGGGCGGTGGCGGACCGCGGCGG - CCGGCGCGCGCGCGCGCGGTGTTGCGGCG 3925507  
QY 477 GATCGTACCGCGGACCGCTCGCGCAGCGCGCGCGG 515  
Db 3925508 CCGCGGTGCGGTGGGGCGGCGACCGACGCGGCGCGCGG 3925546

## RESULT 11

US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: PRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 9.3%; Score 48.6; DB 3; Length 4403765;  
Best Local Similarity 47.7%; Pred. No. 0.17; Indels 1; Gaps 1;  
Matches 173; Conservative 0; Mismatches 189; Indels 1; Gaps 1;  
QY 154 GGGGATGCGGCTCGATTCGAGGTGCGGCGCAGCTCGATCTCGTCCCGCGGTCCC 213  
Db 338841 CGCGCTGCTCGCCCGATCAACAGTCTCTCTGGGGAATACCGGCGCGCGCTGATCG 338782  
QY 214 ACACAGGGGAACCTGGTGGTGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 273  
Db 338781 GCAACGGACCAACGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 338722

QY 274 TGGCGTCACGAGCTCTCGCGGTGGGAGCGGAAACGCGCACGGCTCAGATCCCTGTCA 333  
Db 338721 TGGCAACGCGCGCGCGGAGGATCCGGCGCGCGCGCTCAACGCGCGCGCGCGCA 338662  
QY 334 GTCCGATCGCTCAGTCCGCTGCTCCCTTGGCTTGGAGGATAGCGGTTCACGACA 393  
Db 338661 ACGCGCGCGCGCGCGGCTTCATCGGCAACGCGCGCGCGCGCGCGCGCGCGCA 338602  
QY 394 GCGCACCAACGCGCGCGCGCGCGGCTTCAGCCGATCCGCTCGATGACCAACGCGC 453  
Db 338601 GCACGGGACCG 338543  
QY 454 TGGCGGCTGGGCGGTTCGCGAGATCCGTAACGCGCGGACCGCTTCGCGCGAGCGCG 513  
Db 338542 GCGCGGCTGGCGGTTCGCGCGGATTCGCGCGCGCTTCGCGCGCGCGCGCGCG 338483  
QY 514 GGG 516  
Db 338482 GGG 338480

## RESULT 12

US-09-252-991A-3898/c  
; Sequence 3898, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 3898  
; LENGTH: 606  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-3898

Query Match 9.2%; Score 48; DB 4; Length 606;  
Best Local Similarity 48.2%; Pred. No. 0.21;  
Matches 135; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 1 GGTGACGACGAAACGACCGGTGTCGCCCGCGCGCCCATCAGAACTGCCACTCCGGCG 60  
Db 347 GGGTCACTACCATCAGGTGCGCGCGCGCAACCCGAGCAGGTTGGCGACACTGCG 288  
QY 61 GGGTGACGAGTTCGACAGGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTCGCGCGCA 120  
Db 287 CCTCGTGGCGGCTCGAGGCTGGCGGTGGCTTTCGAGCAGCAGCAACCGGCGCTCG 228  
QY 121 TCCGAAACACCGGCGCAGGATCAGTCAGCAGCAGCGGATGGCGCTTCGATTCGAGCG 180  
Db 227 AGAGCAGCGCGGAGCCAGGACAGCGGTTCGCGTGGCGCGCGCGACAGCGCGCGC 168  
QY 181 TCGCGCGCAGCTCGATCTCGTCGCCCGCGCTCCACACAGCGGAAACTCGGTCGGTGA 240  
Db 167 CGGGACCGACGTCGCTGTCAGCGCCCTCGCGCGCGCGCTTCGATCTCTCCAGCAGGCA 108  
QY 241 GCGGAGCGCCAGCGCGGACAGCTCGTCCAGGCTGGCGTC 280  
Db 107 CGCGCGCGCGCTTAGCGCAGCGCTTCGAGATCGCGCTC 68

## RESULT 13

US-09-417-704-2/c  
; Sequence 2, Application US/09417704  
; Patent No. 6444874

	Query Match	9.0%;	Score 47;	DB 3;	Length 4411529;
	Best Local Similarity	45.3%;	Pred. No. 0.35;		
	Matches 208;	Conservative	0;	Mismatches 250;	Indels 1; Gaps 1;
Qy	57	GGCGGGGTGACAGGTGCGACCAAGTGGAGTTGAGCAGCGAGTCTATGCGCGCTCGGCC	116		
Db	3931301	GGCGTGATCAATGCGGCCACCCAGGCGCTGCTGGGCGGCCGTGTGATCGGTGACGGCGCC	3931360		
Qy	117	GGCATGCCGAAACACCGGGCCAGATACGTGCAGACACGGCGATGCGCGCTCGATCTCG	176		
Db	3931361	AATGCCAGACTCCCGGGGGCCGCGCGGTCTGCTGTTCGGCAACGGCGGG	3931420		
Qy	177	ACGGTCGGCGGCGAGCTCGATCTGTCGCCCGGCTCCACACAGAGGGAACCTGCTCGGT	236		
Db	3931421	GCCGGGCGACGCCGGGCGCCCGGCCAGCGCGCGGCTTGGCGGCGCCGCGATTGTGG	3931480		
Qy	237	GGCAGCGGCGAGCCCAAGCCGGGACAGCTCGTCCAGCTGGCGTCAAGACCTTCGCGGG	296		
Db	3931481	GGCAACGGCGGGCCCGCGCGGCGCGGAGCGGTGGGGGACACGGCGTTCGCGCGGC	3931540		
Qy	297	TCGGGACGGAAACGGCGACGGCTCAGATCCCTGTTCAGTTCGATTCGGCTCAGTCCCGGTC	356		
Db	3931541	GCCGTTGGGTGGGTGTTCCGGGTTGGCGCGCGCGGGGTTCGGTGGGCGCGGTGCGCGC	3931600		
Qy	357	GTCCCCCTTGGCTGGAGGATAGCGGTTCAAGACAGCGGACACACGGCGGGCGGGGGC	416		
Db	3931601	ACCGCGGGGCGGGCGGGCCCGGTGGTTTATCTGGGGCGGGCGGGGCGCGCGTGTCTC	3931660		
Qy	417	GGGGGCGGTTTACGCCGATCCGCTCGATGACACAGCGGCTCGGGGTTCGGGCGGTTCGGCGA	476		
Db	3931661	GGTGGGCGCGTGGCGGACACGGCGGGG-CCGGCGCCCGCGCGAGCTCTGTTCGGCGC	3931719		
Qy	477	GATCCGTACGCCCGGACCGCTTCGCGCACGGCGCGCGG	515		
Db	3931720	CGCGGTGCGGGTGGGGCGGGACACCGACGCGGGGCGCGG	3931758		

QY	417	GGGGGCGGTTCAAGCCGATCCGTTCGATGACCAAGCGGCTCCGGGGTCCGGGCGGTCGGCGCA	476
Db	3931661	GGTGGGGCGGTGGCGGCACCGCGGGG-CGGCGGCGCGCGAGCTCTCTTTCGCGCG	3931719
QY	477	GATCCGTACCGCCGACCGCTCGGCCAGCGCGCCCG	515
Db	3931720	CGCGGTGCGGTGGGGCGGGCACCGACGCGGGCCCCCG	3931758

RESULT 15  
 US-09-833-381-1616/c  
 ; Sequence 1616, Application US/09833381  
 ; Patent No. 6872186  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robison, Keith E.  
 ; TITLE OF INVENTION: No 6672186el Nucleic Acid and Protein Homologs

```

; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1616
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-1616

Query Match          9.0%; Score 46.8; DB 4; Length 1977;
Best Local Similarity 48.0%; Pred. No. 0.37;
Matches 165; Conservative 0; Mismatches 177; Indels 2; Gaps 1;

QY      152  CACGGGGATGCGGCGCTCGATCTCGACGGTTCGGCGCGCAGCTCGATCTCGTCCGCCGGCTC 211
Db      152  CCGCGGAGACGAGCCCGCGCTCTGCTGGGTAGCAGCGCTCGGGCGCGGCTCCACGCT 739

QY      212  CCACACACAGGGGAAACTGGCTCGGTGGCAGCGGCAGCCACCGCGGACAGCTCGTCCAG 271
Db      212  CCGCGGAGACGAGCCCGCGCTCTGCTGGGTAGCAGCGCTCGGGCGCGGCTCCACGCT 739

QY      272  GTCGGGTCACGACCTCTCTCGGCTCGGAGCGGAAACGCGCACCGGTCAGATCCCTGT 331
Db      272  GCGGGATCGCGCGGCGC--CGCGGCGGGCGCGCGGAAACGAAGACGCCGAGCAGCGCG 621

```

QY 332 CAGTCCGATCGCTCAGTCCGCTGCTCCCTTGGCCCTGGGAGGATAGGGTTCCAGAC 391  
Db 620 CGGTTAAAGGGCCAGGACTCTGGCGCCCGCGGGTTGGCGGGGTGAGGGCAACGCTAA 561  
QY 392 GAGCGGCACCAACGCGCGGGCGGGGCGGTTCAGCCGATCCGCTCGATGACCAAGCG 451  
Db 560 GGAACCCCTCAGCGCTCTCGGACTGCGGCGTGTCCCGGCGCCCAAGTTGAAACGCCCG 501  
QY 452 CTGCGGGGTGCGGGCGGTGCGGAGATCCGTACCGCCCGGACC 495  
Db 500 CCAGAGCGCAGAGCGCGCTCGGGAACGTTTGAGACCGTTTC 457

RESULT 16  
US-09-252-991A-180  
; Sequence 180, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 180  
; LENGTH: 1461  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (489)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-180

Query Match 8.7%; Score 45.4; DB 4; Length 1461;  
Best Local Similarity 49.4%; Pred. No. 0.69;  
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 23 GTCCCGCGCGCCCATCAGAACTGCCACTCGCGGGGTGACAGTGCAGCGAGTG 82  
Db 1197 GTCCACCTCGCGGCCAGCAGCCGCGAGCAACGCCGCGAGAGGGCACTCGACCATGAT 1256  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCCGAAACACCGGCCAGGAT 142  
Db 1257 CCCAGTTGAGTTCGCCAGCGGGATTCTCTCGCAGCGCGAGCGCCAGGTACGCGC 1316  
QY 143 CACGTGACGACCGCGATCGCGCTCGATCTGACGGTGGCGCGCAGCTCGATCTCGTC 202  
Db 1317 CTGCGGCCATTCTGTCGAGGTGCGGACCATCGGAAACATCACCGCAGCGCGCTCCCC 1376  
QY 203 GCCCGGCTCCACACCAAGGGAACTGGCTGGTGGCAGCGGAGCCCGCCAGCGGGACA 261  
Db 1377 GCGCGCGGAAACAGCGCGGCACTGGTCTCGAGGATCTGCGTCTGCAGGGTCA 1435

RESULT 17  
US-09-252-991A-166/c  
; Sequence 166, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 166  
; LENGTH: 1560  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-166

Query Match 8.7%; Score 45.4; DB 4; Length 1560;  
Best Local Similarity 49.4%; Pred. No. 0.69;  
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 23 GTCCCGCGCGCCCATCAGAACTGCCACTCGCGGGGTGACAGTGCAGCGAGTG 82  
Db 1188 GTCCACCTCGCGGCCAGCAGCCGCGGCGAGCAACCGCGAGAGGGCACTCGACCATGAT 1129  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCCGAAACACCGGCCAGGAT 142  
Db 1128 CCCAGTTGAGTTCGCCAGCGGGATTCTCTCGGCGAGGCGCAGCGCCAGGTCAAGCGC 1069  
QY 143 CACGTGACGACCGCGATCGCGCTCGATCTGACGGTGGCGCGCAGCTCGATCTCGTC 202  
Db 1068 CTGCGGCCATTCTGTCGAGGTGCGGACCATCGGAAACATCACCGCAGCGCGCTCCCC 1009  
QY 203 GCCCGGCTCCACACCAAGGGAACTGGCTGGTGGCAGCGGAGCCCGCCAGCGGGACA 261  
Db 1008 GCGCGCGGAAACAGCGCGGCACTGGGTCTCGAGGATCTGCGGTCTGCAGGGTCA 950

RESULT 18  
US-09-252-991A-175  
; Sequence 175, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 175  
; LENGTH: 2610  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-175

Query Match 8.7%; Score 45.4; DB 4; Length 2610;  
Best Local Similarity 49.4%; Pred. No. 0.69;  
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 23 GTCCCGCGCGCCCATCAGAACTGCCACTCGCGGGGTGACAGGTGCAGCGAGTG 82  
Db 391 GTCCACCTCGCGGCCAGCAGCCGCGGCGAGCAACCGCGAGAGGGCACTCGACCATGAT 450  
QY 83 GGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGATGCCGAAACACCGGCCAGGAT 142  
Db 451 CCCAGTTGAGTTCGCCAGCGGGATTCTCTCGGCGAGGCGCAGCGCCAGGTCAAGCGC 510  
QY 143 CACGTGACGACCGCGATCGCGCTCGATCTGACGGTGGCGCGCAGCTCGATCTCGTC 202  
Db 511 CTGCGGCCATTCTGTCGAGGTGCGGACCATCGGAAACATCACCGCAGCGCGCTCCCC 570  
QY 203 GCCCGGCTCCACACCAAGGGAACTGGCTGGTGGCAGCGGAGCCCGCCAGCGGGACA 261  
Db 571 GCGCGCGGAAACAGCGCGGCACTGGGTCTCGAGGATCTGCGGTCTGCAGGGTCA 629

RESULT 19  
US-09-773-816-1/c  
; Sequence 1, Application US/09773816  
; Patent No. 6340774  
; GENERAL INFORMATION:  
; APPLICANT: Stanford University  
; APPLICANT: Khosla, Chaitan  
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR  
; TITLE OF INVENTION: ANTAGONISTS  
; FILE REFERENCE: 28600-20210.00  
; CURRENT APPLICATION NUMBER: US/09/773,816  
; CURRENT FILING DATE: 2001-06-08  
; PRIOR APPLICATION NUMBER: US 60/243,458  
; PRIOR FILING DATE: 2000-10-25  
; PRIOR APPLICATION NUMBER: US 60/179,305  
; PRIOR FILING DATE: 2000-01-31  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 23673  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)..(23623)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-773-816-1

Query Match  
Best Local Similarity 49.6%; Pred. No. 0.77;  
Matches 116; Conservative 0; Mismatches 118; Indels 0; Gaps 0;  
QY 56 CGCGGGGTGACAGGTTCGACAGGTGGGAGTTGACAGCCAGCTCATCGCCCTGGCG 115  
DB 12473 CTGCGACCATCATGACCGCTGCCGTGTCAGCCGCTGAAGTCAGCGCGGTG 12414  
QY 116 CGCATGCCGAAACACCGGCCAGGATCAGTCAGACAGCGGATGCGCGCTCGATCTC 175  
DB 12413 CACGGCGCGTAGTGGAGTTGGTATGGTGTGCTGGCGGAAACCGCCCATCAGGCC 12354  
QY 176 GACGGTCGCGCAGCTCCATCTCGTCGCCGCTCCACACAGCGGAACTGCTCGG 235  
DB 12353 GCGCGGTCCAGAGTTTCAGCGCGCGGGTTGATGCGAGCGCGCGGATGCGCGGT 12294  
QY 236 TGGCAGCGGAGCCCGACAGCTCGTCAGGCTGGGCTCAGGACCTC 289  
DB 12293 GCGCTCCGCGCAGCTCCAGGATCAGACGTCGCGCGCGCCAGCGCGCAGCTC 12240

RESULT 20  
US-09-252-991A-6994/c  
; Sequence 6994, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6994  
; LENGTH: 1125  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6994

Query Match  
8.6%; Score 44.8; DB 4; Length 1125;

Best Local Similarity 48.4%; Pred. No. 0.9;  
Matches 155; Conservative 0; Mismatches 162; Indels 3; Gaps 1;  
QY 4 GCAGCAGAACGACCGGTGTCGCCCCGCGGCCCATCAGAACTGCCATCTCGCGGGG 63  
DB 767 GCGCCACCATCAGCAGGTCTCCAGGCCGATGCCATGCCGCGCTCAGCGGGC 708  
QY 64 TGACCAAGTTCGACAGGTGGAGTTGACAGCAGCAGTTCATCGCGCGCTCGCGCGCATGC 123  
DB 707 TGCTCAGCGCGACCGCACCGGTACAGCAGCAAGCCAT--CCACAGCGCCAGCCACT 651  
QY 124 CGAAACACCGCGCCAGGATCAGTGTGAGCAGCGGCGATGCCGCTCGATCTCAGCGTGC 183  
DB 650 CGAAAGTCGGCGGATGCGCGCAACAGCAGCAGCGCGCGCGAGATCAGGATCGCCA 591  
QY 184 GCCGAGCTCGATCTCGTCGCCCGGCTCCACACAGCGGAAACTGGCTCGGTGGCAGCG 243  
DB 590 GCAGCAGCGCGCGCAGGAAGTTCTGCGCCCGCAGCAGCGGAGGAGTGCCTATCA 531  
QY 244 GCAGCCCGCAGCGGACAGTCTGCTCAGGCTGCGGTACCGACCTTCGCGGGTTCGGAG 303  
DB 530 GCGAGCATCATCAGCGCGACCGAGCAGGCTGCTGAGGGCGCGCTTCAGGCGCTGCTGA 471  
QY 304 CGAAACCGCGCACCGCTCAG 323  
DB 470 TCCACAGCGCGCGCAG 451

RESULT 21  
US-09-252-991A-7017/c  
; Sequence 7017, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7017  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7017

Query Match  
Best Local Similarity 48.4%; Pred. No. 0.91;  
Matches 155; Conservative 0; Mismatches 162; Indels 3; Gaps 1;  
QY 4 GCAGCAGAACGACCGGTGTCGCCCCGCGGCCCATCAGAACTGCCATCTCGCGGGG 63  
DB 1252 GCGCCACCATCAGCAGGTCTCCAGGCCGATGCCATGCCGCGCTCAGCGGGC 1193  
QY 64 TGACCAAGTTCGACAGGTGGAGTTGACAGCAGCAGTTCATCGCGCGCTCGCGCGCATGC 123  
DB 1192 TGCTCAGCGCGACCGCACCGGTACAGCAGCAAGCCAT--CCACAGCGCCAGCCACT 1136  
QY 124 CGAAACACCGCGCCAGGATCAGTGTGAGCAGCGGCGATGCCGCTCGATCTCAGCGTGC 183  
DB 1135 CGAAAGTCGGCGGATGCGCGCAACAGCAGCAGCGCGCGCGAGATCAGGATCGCCA 1076  
QY 184 GCCGAGCTCGATCTCGTCGCCCGGCTCCACACAGCGGAAACTGGCTCGGTGGCAGCG 243  
DB 1075 GCAGCAGCGCGCGCAGGAAGTTCTGCGCCCGCAGCAGCGGAGGAGTGCCTATCA 1016  
QY 244 GCAGCCCGCAGCGGACAGTCTGCTCAGGCTGCGGTACCGACCTTCGCGGGTTCGGAG 303  
DB 1015 GCGAGCATCATCAGCGCGACCGAGCAGGCTGCTGAGGGCGCGCTTCAGGCGCTGCTGA 956

QY 304 CGGAACGCGCAGCGCTCAG 323  
Db 955 TCCACAGCAGCGCGCGCAG 936

## RESULT 22

US-09-252-991A-7111  
; Sequence 7111, Application US/09252991A  
; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 7111

; LENGTH: 1947

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-7111

Query Match 8.6%; Score 44.8; DB 4; Length 1947;  
Best Local Similarity 48.4%; Pred. No. 0.91;  
Matches 155; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 4 CGACAGCAAGACCGCTGCTGCGCCCGCGCCATCAAGAACTGCGCACTCCGCGCGGG 63  
Db 636 GCGCCACCATCAGCAAGTCTCCAGCGCATGCCATCGCGCTCAGCGGCG 755  
QY 64 TGACAGGTGACAGAGTGGAGTTGAGCAGCAGCTCATCGCGCTGCGCGCGCATGC 123  
Db 756 TGCTCAGCGCAGCGCGACCGACGACGACGACGACGACGACGACGACGACGAC 812  
QY 124 CGAAGACCGCGCGAGGATCAGTGCAGCAGCGGATCGCGCTCGATCTCGAGGTG 183  
Db 813 CGAAGTGGCGGATCGCGGACGACGACGACGACGACGACGACGACGACGACGAC 872  
QY 184 GCGCAGGTGATCTGCTGCTGCGCGCGCTCCACACAGCGGGAATCTGCTCGGTGCG 243  
Db 873 GCAGCAGCGCGCGAGGAGTCTGCGCGCGCAGCAGCGCGGGAAGTCCCATCA 932  
QY 244 GCAGCGCGCGCGGACAGCTGCTCCAGCTGCGGTACGAGACCTCTCGCGGCTCGGAG 303  
Db 933 GCGAGCAGATCAGCGCGCGCGGACGAGGCTGCTGAGGCGCGCTTCCAGCGCTGTGA 992  
QY 304 CGGAACGCGCAGCGCTCAG 323  
Db 993 TCCACAGCAGCGCGCGCAG 1012

## RESULT 23

US-09-616-289-45/c

; Sequence 45, Application US/09616289  
; Patent No. 6632923

## GENERAL INFORMATION:

; APPLICANT: Lees, Am M.

; APPLICANT: Lees, Robert S.

; APPLICANT: Law, Simon W.

; APPLICANT: Arjona, Anibal A.

; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING

; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING

; FILE REFERENCE: 10797-004001

; CURRENT APPLICATION NUMBER: US/09/616,289

; CURRENT FILING DATE: 2000-07-14

; PRIOR APPLICATION NUMBER: US 09/517,849

; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; PRIOR FILING DATE: 1997-06-03  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 45  
; LENGTH: 1614  
; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1614)

US-09-616-289-45

Query Match 8.6%; Score 44.6; DB 4; Length 1614;  
Best Local Similarity 47.4%; Pred. No. 0.99;  
Matches 172; Conservative 0; Mismatches 184; Indels 7; Gaps 1;

QY 154 CGCGATGCGCGCTCGATCTCGACCGTCCGCGCGCAGCTCGATCTCGTCCGCCCGGCTCCC 213  
Db 727 CGSCCGCACCAGCGCCCTCCGCGCGGCTGTGGCTGCGCGCGCGCGCTGTGCT 668  
QY 214 ACACAGGGGAACTGGCTCGGTGGCAGCGGAGCCCGCAGCCCGGACAGCTGTCCAGGC 273  
Db 667 GCTGTGCGCGCGCGCGCGCTGTGGCGCGCGCGCGCGCTCCCGCGCGCGCGCG 608  
QY 274 TGGCGTCAAGACCTCTCGCGGTCCGGAGCGGAAACGCGCACGGCTCAGATCCCTGTCA 333  
Db 607 CGSCCG 548  
QY 334 GTGCGATCGGCTAGTTCGCGCTCGTCCCTTGGGCTGGAGAGTAGCGGTTACAGACA 393  
Db 547 CG 495  
QY 394 GCGGACACCG 453  
Db 494 GCGGCG 435  
QY 454 TCGCGGCTCGGCGCGCTCGCGCGAGATCCGTACCGCGCGCGCGCGCGCGCGCGCGCG 513  
Db 434 GCGGCG 375  
QY 514 GCG 516  
Db 374 GCG 372

## RESULT 24

US-09-252-991A-1273

; Sequence 1273, Application US/09252991A

; Patent No. 6551795

## GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 1273

; LENGTH: 1839

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-1273



Query Match 8.6%; Score 44.6; DB 4; Length 1839;  
Best Local Similarity 49.0%; Pred. No. 0.99;  
Matches 119; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 15 GACCGTGGTGGCCCGCCGATCAGCAAGTGCACATCCGCGGGGTGACCAAGTGG 74  
|||  
Db 1496 GAGAGGAAGTGGTGGCCCGCCGCTCCAGGCTTCACACAGGCTCGGTTGGTGGG 1555  
|||

QY 75 ACCAGTGGAGTGGAGCAGCAGTCAATCGCCGCTGCGCGGATGCGGAAACACCGG 134  
|||  
Db 1556 GAGGTGAGGAAGATGATCGGCACCAACTCTCTCGCGGCTGTGCTTGATCGCGCGGCA 1615  
|||

QY 135 GCACGATACATGTCAGCAGCGGATGCGGCTCGATCTCGAGGTGCGCGCAGTGG 194  
|||  
Db 1616 GCTCGAGCATCATCATCCGCGCATCATCATCTCATGAGACAGTTGCGCGCTGG 1675  
|||

QY 195 ATCTGTCGCCCGGCTCCACACAGGGAATGCTCGGTGGCAGCGGACCGCCAGC 254  
|||  
Db 1676 GTCTGAAACAGCGCCACCGCTCCAGGCGGTTGCGGCGGTGAGCAGCGGATGCCCTGG 1735  
|||

QY 255 CGG 257  
|||  
Db 1736 CGG 1738  
|||

## RESULT 25

US-09-252-991A-1333/c  
; Sequence 1333, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCES: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 1333  
; LENGTH: 1980  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-1333

Query Match 8.6%; Score 44.6; DB 4; Length 1980;  
Best Local Similarity 49.0%; Pred. No. 0.99;  
Matches 119; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 15 GACCGTGGTGGCCCGCCGATCAGCAAGTGCACATCCGCGGGGTGACCAAGTGG 74  
|||  
Db 599 GAGAGGAAGTGGTGGCCCGCCGCTCCAGGCTTCGACACGCGCTCGGTTGGTGGG 540  
|||

QY 75 ACCAGTGGAGTGGAGCAGCAGTCAATCGCGGCTGCGCGGATGCGGAAACACCGG 134  
|||  
Db 539 GAGGTGAGGAAGATGATCGSCACCAACTCTCTCGCGGCTGTGCTTGATCGCGGCA 480  
|||

QY 135 GCAAGATCATGTCAGCAGCAGGATGCGGCTTCATCTGAGGTGCGGCGCAGTGG 194  
|||  
Db 479 GCTCGAAGCATTCATCAGCGCATCATCATGCTCATGAGACCAAGTTGCGGCGCTGG 420  
|||

QY 195 ATCTGTCGCCCGGCTCCACACAGGGAATGCTCGGTGGCAGCGGACCGCCAGC 254  
|||  
Db 419 GTCTGAAACAGCGCCACCGCTCCAGGCGGTTGCGGCGGTGAGCAGCGATGCCCTGG 360  
|||

QY 255 CGG 257  
|||  
Db 359 CGG 357  
|||

## RESULT 26

US-09-616-289-50/c  
; Sequence 50, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.

US-09-616-289-48/c  
; Sequence 48, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.  
; APPLICANT: Law, Simon W.  
; APPLICANT: Arjona, Anibal A.  
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
; TITLE OF INVENTION: ATHEROSCLEROSIS  
; FILE REFERENCE: 10797-004001  
; CURRENT APPLICATION NUMBER: US/09/616,289  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 09/517,849  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 08/979,608  
; PRIOR FILING DATE: 1997-11-26  
; PRIOR APPLICATION NUMBER: US 60/031,930  
; PRIOR FILING DATE: 1996-11-27  
; PRIOR APPLICATION NUMBER: US 60/048,547  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 2561  
; TYPE: DNA  
; ORGANISM: Oryctolagus cuniculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (246)...(1895)  
US-09-616-289-48

Query Match 8.6%; Score 44.6; DB 4; Length 2561;  
Best Local Similarity 44.7%; Pred. No. 0.99;  
Matches 173; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

QY 130 ACCGGGCGAGATCAGTGCAGCAGCGCGATGCGCGCTTCGATTCGACGTCGCGCA 189  
|||  
Db 507 AGCTGACCGGAGCAGCGCGCTGCTGGATCAGTTTCTCGAGCTCGCGCGCTGGCT 448  
|||

QY 190 GCTCGATCTGTCGCCCGGCTCCACACAGGGAATCGCTCGTGGCAGCGGCGAGCC 249  
|||  
Db 447 CCGGCTCGGCGCGTGGCGCGCGCGCACCATCCGCGAGATGCGTCCAGTCCGCGCG 388  
|||

QY 250 CCAGCGGGGACAGTCTGTCAGGCTGCGCTCAGGACCTTCGCGGGTCCGAGCGGAAA 309  
|||  
Db 387 CTTTGGCGGAGCGCAGGATGCGATGCTGTCAGATCCACTCTTGTAGTGGCGGGAAG 328  
|||

QY 310 CGCGCAGGCTCAGATCCCTCTCAGTCGATCGATCGGCTCAGTCCGCTGCTCCCTTGGCC 369  
|||  
Db 327 CGGCGGACGACGAGCGCGCGCGCGCGCTGCTGCGCGCGCGCGCTCTCGGCGCGGG 268  
|||

QY 370 TGGGAGATAGCGGTTTCAACGACGCGGACACGCGGGCGGGCGGGCGGGGTTTCAG 429  
|||  
Db 267 GTAGGCGCGGGGCGCGCGCGCTCCCTCCCGCGCGCGCTGCACGTCGCGCTCCCT 208  
|||

QY 430 CCGATCCGCTCGATGACAGCGGCTGCGGGTTCGGGCGGTTCGCGAGATCCGTACGCC 489  
|||  
Db 207 CCCTCCAGCGCGGCGCGCGCTCCCTCCCGGCTCCCTCCCTCCCGCGCTGCGCGCTCCTC 148  
|||

QY 490 CGGACCGGCTCGGCGCAGCGCGCGCGG 516  
|||  
Db 147 CGCTCCCGCGCGCGCGCGCGCGCGG 121  
|||

## RESULT 27

US-09-616-289-50/c  
; Sequence 50, Application US/09616289  
; Patent No. 6632923  
; GENERAL INFORMATION:  
; APPLICANT: Lees, Ann M.  
; APPLICANT: Lees, Robert S.



```

; APPLICANT: Law, Simon W.
; APPLICANT: Arjona, Anibal A.
; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING
; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING
; TITLE OF INVENTION: ATHEROSCLEROSIS
; FILE REFERENCE: 10797-004001
; CURRENT APPLICATION NUMBER: US/09/616,289
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 09/517,849
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 08/979,608
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 60/031,930
; PRIOR FILING DATE: 1996-11-27
; PRIOR APPLICATION NUMBER: US 60/048,547
; PRIOR FILING DATE: 1997-06-03
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 12425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-616-289-50

Query Match      8.6%; Score 44.6; DB 4; Length 12425;
Best Local Similarity 47.4%; Pred. No. 1;
Matches 172; Conservative 0; Mismatches 184; Indels 7; Gaps 1;

Qy 154 CGGCGATGCGCGCTCGATCTCGACGGTCGGCGCGCAGCTCGATCTCGTCGCCGGGTCCC 213
Db 3558 CGGCGCGCAGCGCGCCCCCTCGCGGCGCGCTGTGCTCGCGGCGGGGTGCTGCT 3499

Qy 214 ACACGAGGGAATGCTGCTGGTGGCAGCGGACGCCACCGGGACAGCTGTCAGGC 273
Db 3498 GCTGTGCGCGCGCGCGGCTGTGTGCGCGCGCGCAGCGCGGCTCCGGCGCGCAGCG 3439

Qy 274 TGGCGTCACGACCTCTCGCGGGTCGGGAGCGGAACGCGACGGCTCAGATCCCTGTCA 333
Db 3438 CGGCGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3379

Qy 334 GTGCAATCGCTAGTCCGGTCTGTCCTTGGCTTGGAGGATAGCGGTTTACGACGA 393
Db 3378 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3326

Qy 394 GCGGCACACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 453
Db 3325 GCGGGGCGACGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3266

Qy 454 TGGCGGGTCTGGGGCGCTGGCGAGATCCGTACCGCCCGGACCGGCTCGGCGACGCGCG 513
Db 3265 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTGGTGGCGGGCGGGCGCTGGCGCGCG 3206

Qy 514 GGG 516
Db 3205 GCG 3203

```

```

RESULT 28
US-09-410-551B-1
; Sequence 1, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; CURRENT FILING DATE: 1999-10-01

```

```

; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 77536
; TYPE: DNA
; ORGANISM: Streptomyces hygroscopicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (52275)...(71465)
US-09-410-551B-1

Query Match      8.6%; Score 44.6; DB 4; Length 77536;
Best Local Similarity 44.7%; Pred. No. 1;
Matches 173; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Qy 10 CGAACGACCGGTGGTTCGCCCGCGCGCCCATCACGAATGCCACTCCCGCGGGGTGACCA 69
Db 17190 CCATCGCTCCCGCGCGGCAGTCGCGCGGGATGACCTGACTGCGCAATGCCACCGC 17249

Qy 70 GGTGAGACAGTGGAGTTGACAGCAGCTCATCGCGCCTGCGCGGCATCGGAAC 129
Db 17250 GGGCGCGCTCTCGAGGCTGAGGGGTTCGGGCCACGACGCGCGCGCATTCGCCCTGGG 17309

Qy 130 ACCGGGCCAGAGTACAGTGCAGCAGCGGGATGCGCGCCTCGATCTCGACGAGTCGCGCGCA 189
Db 17310 AGTGTCGATCAACCGGTTCGGCACGACCCCATGCGCTGCCACAGCGCGCCAGGCTCA 17369

Qy 190 GCTCGATCTGTCGCCGCTCCACACACGAGGGAACTGGCTGGTGGCAGCGCAGCC 249
Db 17370 CCGCGACCGCCAGCTGGCGCGCTGGACCACTCCACCCGCTCGCGCCACATCCCGGCGCG 17429

Qy 250 CCAGCGGGACAGCTGTTCCAGGCTGGGCTCAGCGACCTCTCGCGGTCGGGAGCGGAAA 309
Db 17430 CCAACATCTCCGCGCATCCAGCGCCGTGTGCGGAGCAACGCCGTAGAGCGCACTCTCCA 17489

Qy 310 CGCGCAGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCCGGTCGTCCCTTGGCC 369
Db 17490 TAGCGCGCGGAACACCGCGAGTGGGCAATGAGTTCCACGCCCATCGGCACTGGG 17549

Qy 370 TGGAGGATAGCGGTTACGACGAGCG 396
Db 17550 CGCCCTGGCGCGGAAGACGAACACCG 17576

```

```

RESULT 29
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;
; GENERAL INFORMATION:
;
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
;
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
;
; NUMBER OF SEQUENCES: 15
;
; CORRESPONDENCE ADDRESSES:
;
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
;
; COUNTRY: USA
;
; ZIP: 46285
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM Compatible
;
; OPERATING SYSTEM: MS-DOS
;
; SOFTWARE: ASCII(DOS) Text only

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,764
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
;
; US-08-804-227C-7
;
; Query Match 8.5%; Score 44.4; DB 2; Length 44377;
; Best Local Similarity 44.6%; Pred. No. 1.1;
; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
;
; QY 8 CACGACGACGCGTGTGTCGCCCGCGCCGATCAGCACTCCCACTCCGGCGGGGTGAC 67
; DB 40450 CCGGCTCCACCCGAGACGGCGCTACCGCCCTGCACAGGCATCGAGCGGGAAGC 40509
;
; QY 68 CAGGTGACAGGTGGAGTTGAGAGCCAGCTCATCGCGGCTCGCGCGGATGCCGA 127
; DB 40510 CACGGTCACCGTGGCGGACATCGACTGGGAACGGTTCCGCCCGGCTTACCGCCTTCCG 40569
;
; QY 128 ACACCGGGCCAGGATCACGTGACGACGGCGCTCGCGCTCGATCTCGACGCTCGGCG 187
; DB 40570 TCCAGCCCCCTGATCGCGGCTACCGCGGCGCTACGGCGCGCGCGCGCGCG 40629
;
; QY 188 CAGCTGATCTGTCGCCGGGTCCACACACCGGGGAAACTGGCTCGGTGGAGCGGCG 247
; DB 40630 CGCCGAGGACACCCCGCCCGGCTCTCGCGGCGCGCGCGCGCGCGCGCG 40689
;
; QY 248 CCCAGCGCGGACAGCTCGTCCAGGCTGCGGTGCGGTGCGGCTCTCGCGGCGCGG 307
;
; Query Match 8.5%; Score 44.4; DB 2; Length 44377;
; Best Local Similarity 44.6%; Pred. No. 1.1;
; Matches 174; Conservative 0; Mismatches 216; Indels 0; Gaps 0;
;
; QY 8 CACGACGACGCGTGTGTCGCCCGCGCCGATCAGCACTCCCACTCCGGCGGGGTGAC 67
; DB 40450 CCGGCTCCACCCGAGACGGCGCTACCGCCCTGCACAGGCATCGAGCGGGAAGC 40509
;
; QY 68 CAGGTGACAGGTGGAGTTGAGAGCCAGCTCATCGCGGCTCGCGCGGATGCCGA 127
; DB 40510 CACGGTCACCGTGGCGGACATCGACTGGGAACGGTTCCGCCCGGCTTACCGCCTTCCG 40569
;
; QY 128 ACACCGGGCCAGGATCACGTGACGACGGCGCTCGCGCTCGATCTCGACGCTCGGCG 187
; DB 40570 TCCAGCCCCCTGATCGCGGCTACCGCGGCGCTACGGCGCGCGCGCGCGCG 40629
;
; QY 188 CAGCTGATCTGTCGCCGGGTCCACACACCGGGGAAACTGGCTCGGTGGAGCGGCG 247
; DB 40630 CGCCGAGGACACCCCGCCCGGCTCTCGCGGCGCGCGCGCGCGCGCGCG 40689
;
; QY 248 CCCAGCGGACAGCTCGTCCAGGCTGCGGTGCGGTGCGGCTCTCGCGGCGCGG 307
; DB 40690 GTTCGCCCTTGGACCTGCTGCTCGGCGCTGCTGCGGCGCTTCCGCGGCTTCCGAGAGCG 40749
;
; QY 308 AACCGGACAGGCTCAGATCCCTGTGATGCGCATCGGCTCAGTGCAGGCTGCTCCCGCTTGG 367
; DB 40750 CCGGGTCGACGCGCGCGCGCCCTTCCGGAACCTCGGCTTCCGCTGCTGCTCGCGGCTGCG 40809
;
; QY 368 CTGGGAGATAGCGTTACACAGCGG 397
; DB 40810 GTTCGCCCGCGGCTGCGCGGAGACACCGG 40839
;
; RESULT 30
; US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.

```

Db 40690 GCTGCCCTGAGCTGCTCCGCGCAGCTGCGCGGCTCTCGGCCCACTCCGAGGACGC 40749  
QY 308 AACGGCAGCGCTCAGATCCCTCTCAGTCGATCGGCTCAGTCGCGGTCTGTCCTCCCTGG 367  
Db 40750 CCGGCTCAGCGCGCGCGCCCTCCGCGGACTCGGCTTCGACTCGCTCGCGCGGTGCG 40809  
QY 368 CTGGAGGATAGCGGTTACGACGAGCGG 397  
Db 40810 CTGCGCGCGCGGTGGCGCGAGACACCGG 40839

## RESULT 31

US-09-252-991A-13718/c  
; Sequence 1, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13718  
; LENGTH: 927  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13718

Query Match 8.5%; Score 44.2; DB 4; Length 927;

Best Local Similarity 48.6%; Pred. No. 1.2; Mismatches 128; Indels 0; Gaps 0;  
Matches 121; Conservative 0;

QY 32 GCCGCCCATCAGCAACTGCCACTCCGCGGGGTTGACAGGTCCACACGAGTGGCGAGTTGAG 91  
Db 384 GCGCGCATCACTCCCGATGCTGCTCGGTGAATCATCATCGATCGGCGCGCTGGTC 325  
QY 92 CAGCAGTCTATCCGCGCTCGCGCGGATGCGGAAACACCGGCGCAGGATCACTGTCAG 151  
Db 324 GCGGCTCAGCAGGAGGAGCTGCTCCATCAGCTCGCGCGCGGAGACTCCGCTGTCATCAG 265  
QY 152 CACGCGATGCGCGCTCGATCTCGACGTCGCGCGGCGAGCTCGATCTCGTCGCGCGCTC 211  
Db 264 GATCAGTCGCGGACGAGCGCTTCGACGCGCGCGGCGAGTCTGATGTGAGCCCTGACTC 205  
QY 212 CCACACAGGGAATGCTGCTGGTGGCGAGCGGCGAGCCCGGAGACAGCTCGTCCAG 271  
Db 204 GTCCACCACTCGAAACCGCTCTCGAGCAGGCGGCGGCGCTCGAGGCTCTTCTTGG 145  
QY 272 GCTGGCGTC 280  
Db 144 GGTGTGTC 136

## RESULT 32

US-09-105-537-1/c  
; Sequence 1, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438U51  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43

Query Match 8.5%; Score 44.2; DB 4; Length 15872;

Best Local Similarity 45.8%; Pred. No. 1.2;  
Matches 189; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

QY 32 GCGGCCCATCAGAACTGCCACTCCGCGGGGTGACCAAGTCCGACGAGTGGAGTTGAG 91

; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

## Query Match

Best Local Similarity 45.8%; Score 44.2; DB 3; Length 15872;

Matches 189; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

QY 32 GCGGCCCATCAGAACTGCCACTCCGCGGGGTGACCAAGTCCGACGAGTGGAGTTGAG 91  
Db 2191 GCGGTGCGAGACGCTGAGCCGCTTCGTCGCGGGGCTGTTCCGCGAGCAGCTCGCGAC 2132  
QY 92 CAGCCAGCTCATCGCGCTCGCGGATCGCGGATCCGAAACACACCGGCGCAGATCACGTGAG 151  
Db 2131 GTCGAGGAGCGCGCTCGCGCCCGCGGAGGACGACCGCGCGGGGCGCTTCCGCGCGC 2072  
QY 152 CAGCGGATGCGCGCTCGANCTCGACGTCGCGCGCAGCTCGATCTCGTCCCGCGGTC 211  
Db 2071 GAGGAGAGCTCCCGCTCGCGTCCGCGAGGTGCGAAAGCACTTCTCTCGCTCGCTC 2012  
QY 212 CCACACCGGGAATACTGCTCGGTGGCGGCGAGCGGCGGCGGCGGCGGCGGCGGCGG 271  
Db 2011 GACCGGACCATCGCGCGCGCTCGGAGAGCGCTGATGAGCGCGCGCGCGCGCGCGAC 1952  
QY 272 GCTGCGTACGAGACTCTCGCGGTGCGGAGCGGAAACGCGCACGCGTTCAGATCCCTGT 331  
Db 1951 GAGCGCGCGCGCTCGCGCAGCAGGACCGCGCGCGGCGGCGGCGGCGGCTGATCTC 1893  
QY 332 CAGTCGATCGCTCAGTCGCGGTGCTCCCTTGGCTGGGAGGATACCGGTTCCAGAC 391  
Db 1892 CGACGAGTGGCGCGGAGCAGTCCGCGTGAACCGCCCGAGACTCGACGAGCGGTGGA 1833  
QY 392 GAGCGGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444  
Db 1832 GGGCGACCTCCACGCGGAGAGCGGCGGCTGTGTGACGGTGGGTCGAGG 1780

## RESULT 33

US-09-091-609-1/c  
; Sequence 1, Application US/09091609  
; Patent No. 6600029  
; GENERAL INFORMATION:  
; APPLICANT: SHERMAN, DAVID H.  
; APPLICANT: WILLIAMS, MARK D.  
; APPLICANT: XUE, YONGQUAN  
; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
; FILE REFERENCE: 600.297US2  
; CURRENT APPLICATION NUMBER: US/09/091,609  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: PCT/US96/20119  
; EARLIER FILING DATE: 1996-12-18  
; EARLIER APPLICATION NUMBER: 60/008,847  
; EARLIER FILING DATE: 1995-12-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (20)....(13909)  
US-09-091-609-1



QY 214 ACACAGGGGAACTGGCTCGTGGCAGCGGCGCCAGCGGAGCAGCTCGTCAGGC 273  
DB 3562 ATGGCGGCGCGGAGTGTGATGGGTGGCGCGCGGACCTGCTGTGAGCGACGCCGCTC 3503  
QY 274 TGGGCTCACGACCTCTCGCGGGTGGGAGCGGAAACGCGCACGCTCAGATCCCTGTCA 333  
DB 3502 GGGGCTCGGGGCCAGCGGCTCGGAGCTCGGCGCGCGCTCGGACGCTGGGCGCTCAG 3443  
QY 334 GTGCGATCGGCTCAGTGGCGGTCTGCTCCCTTGGCTGGGAGGATAGGGTTACGACGA 393  
DB 3442 TGGGCGCCCGCGCGGCT 3383  
QY 394 CGCGCACACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 453  
DB 3382 GGGGGGGCGGGGTGACGGCTCGCGCCCGCTTGGCGACAGCGCGGCGGCGGCGGCGAT 3323  
QY 454 TGGGGGTGGGGCGGTGGCGAGATCGTACCGCC 489  
DB 3322 GAGGACTGGGATCGGAGGCTTACCGCTCGCTCC 3287

## RESULT 36

US-08-658-136-1/c  
; Sequence 1, Application US/08658136  
; Patent No. 6071717  
; GENERAL INFORMATION:  
; APPLICANT: KLINGER, KATHERINE W  
; APPLICANT: LANDES, GREGORY M  
; APPLICANT: BURN, TIMOTHY C  
; APPLICANT: CONNORS, TIMOTHY D  
; APPLICANT: DACKOWSKI, WILLIAM  
; APPLICANT: GERMINO, GREGORY  
; APPLICANT: QIAN, FENG  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENZYME CORPORATION  
; STREET: ONE MOUNTAIN ROAD  
; CITY: FRAMINGHAM  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 01701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: LASSEN, ELIZABETH  
; REGISTRATION NUMBER: 31,845  
; REFERENCE/DOCKET NUMBER: GEN4-17.8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 508-872-8400  
; TELEFAX: 508-872-5415  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 53577 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-658-136-1

Query Match 8.5%; Score 44; DB 3; Length 53577;  
Best Local Similarity 44.4%; Pred. No. 1.3;  
Matches 176; Conservative 0; Mismatches 220; Indels 0; Gaps 0;  
QY 94 GCCAGCTCATCGCGCTGTGGCGGATGCGGAAACACCGGCGCAGGATCAGGTGACGA 153

DB 3682 GCCAGGCGCAGCGGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3623  
QY 154 CGGCGATGCGCGCTCGATCTCGACGGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 213  
DB 3622 CGCGGTCCCGAGCGCGCGCGCGGAGGCGGAGGCTCAGGCGGCGGCGGCGGCGGCGGCGG 3563  
QY 214 ACACAGGGGAACTGGCTCGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 273  
DB 3562 ATGGCGGCGCGGCGGCTGATGGGTGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3503  
QY 274 TGGCGTACGGAACCTCTCGCGGGTGGGAGCGGAAACGCGCACGCTCAGATCCCTGTCA 333  
DB 3502 GGGGCTCGGGGCCAGCGGCTCGGAGCTCGGCGCGCGCTCGGACGCTGGGCGCTCAG 3443  
QY 334 GTGCGATCGGCTCAGTGGCGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 393  
DB 3442 TGGGCGCCCGCGCGGCT 3383  
QY 394 CGCGCACACAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 453  
DB 3382 GGGGGGGCGGGGTGACGGCTCGCGCCCGCTTGGCGACAGCGCGGCGGCGGCGGCGAT 3323  
QY 454 TGGGGGTGGGGCGGTGGCGAGATCGTACCGCC 489  
DB 3322 GAGGACTGGGATCGGAGGCTTACCGCTCGCTCC 3287

## RESULT 37

US-09-252-991A-808/c  
; Sequence 808, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107195.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 808  
; LENGTH: 1197  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-808

Query Match 8.4%; Score 43.8; DB 4; Length 1197;  
Best Local Similarity 46.8%; Pred. No. 1.4;  
Matches 138; Conservative 0; Mismatches 157; Indels 0; Gaps 0;  
QY 113 CGCGCGATGCCGAAACACCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 172  
DB 1191 CGCGCGCAGAGGTCGAGCGGCTCACATGCTCGACCTGCCAGCGCGGCGGCGGCGGCGG 1132  
QY 173 CTCGAGGTGGCGCGCGCTCGATCTCGTGGCGCGGCTCCACACAGCGGCGGCGGCGGCGG 232  
DB 1131 ACCGAGCTCAGGTCGCGGTCCAGCTCGCGGATCGGCTGCTTGAGGAGAAAGTCGCGG 1072  
QY 233 CGGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 292  
DB 1071 GCGGCAAACTCCAGCGGCGACACGCGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1012  
QY 293 CGGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 352  
DB 1011 GCGGAGAAACGGTTGACGCAAGACTCGCTGGCAACCGTTTCAGGCGCACCGCTCGGGG 952  
QY 353 GGTCTGTCCCGCTTGGCTGGGAGGATAGCGGTTACGACGAGCGGCGGCGGCGGCGGCGG 407  
DB 951 GATGACCAAGGTTACCGGCTCGCGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 897

```
RESULT 38
US-09-105-537-3/c
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-3
Query Match      8.4%; Score 43.6; DB 3; Length 13613;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 314 CACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCGGGTCGTCGCCCTTGGCCCTGGG 373
DB 3473 CAGGGGTGAGGCTCGGGGCGAGGGGTGCGGCTCACTGCGGGGTGCGCCGGCGGCGGC 3414
QY 374 AGGATAGCGGTTACAGACGAGCGGACACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433
DB 3413 GGTGCTGCGCGCGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGGTGGGGTCCGCGA 3354
QY 434 TCCGCTCGATGACACGAGCGGCTCGGGGTCGGGGCGGTGCGGACGATCCGTACCGCCCGGA 493
DB 3353 AGGTCTCTCGCGGACGCGGTGCGGGCGGTGGCGACGAGGGGTGCTCGAGGATGCGGA 3294
QY 494 CCGCTCGCGGACGCGCGCGG 515
DB 3293 CGACGGGCTCCGCGACGGGCTG 3272

RESULT 39
US-09-320-878-19/c
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
```

```
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19
Query Match      8.4%; Score 43.6; DB 3; Length 38506;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 314 CACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCGGGTCGTCGCCCTTGGCCCTGGG 373
DB 37470 CAGGGGTGAGGCTCGGGGCGAGGGGTGCGGCTCACTGCGGGGTGCGCCGGCGGCGGC 37411
QY 374 AGGATAGCGGTTACAGACGAGCGGACACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433
DB 37410 GGTGCTGCGCGGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGGTGGGGTCCGCCGA 37351
QY 434 TCCGCTCGATGACACGAGCGGCTCGGGGTCGGGGCGGTGCGGACGATCCGTACCGCCCGGA 493
DB 37350 AGGTCTCTCGCGGACGCGGTGCGGGCGGTGGCGACGAGGGGTGCTCGAGGATGCGGA 37291
QY 494 CCGCTCGCGGACGCGCGCGG 515
DB 37290 CGACGGGCTCCGCGACGGGCTG 37269

RESULT 40
US-09-141-908-1/c
; Sequence 1, Application US/09141908
; Patent No. 6503741
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
; TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
; FILE REFERENCE: 300622002100
; CURRENT APPLICATION NUMBER: US/09/141,908
; CURRENT FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: PROV. 60/076,919
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: PROV. 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-141-908-1
Query Match      8.4%; Score 43.6; DB 4; Length 38506;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 103; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 314 CACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCGGGTCGTCGCCCTTGGCCCTGGG 373
DB 37470 CAGGGGTGAGGCTCGGGGCGAGGGGTGCGGCTCACTGCGGGGTGCGCCGGCGGCGGC 37411
QY 374 AGGATAGCGGTTACAGACGAGCGGACACGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433
DB 37410 GGTGCTGCGCGGAGCGGCTCCAGCTCGGGGACGATCCCGCGGGGGTGGGGTCCGCCGA 37351
QY 434 TCCGCTCGATGACACGAGCGGCTCGGGGTCGGGGCGGTGCGGACGATCCGTACCGCCCGGA 493
DB 37350 AGGTCTCTCGCGGACGCGGTGCGGGCGGTGGCGACGAGGGGTGCTCGAGGATGCGGA 37291
```

QY 494 CCGCTCGGCGGCGGCGGCGG 515  
Db 37290 CGAGGCGTCCCGCAGCGGCTG 37269

## RESULT 41

US-09-657-440-19/c  
; Sequence 19, Application US/09657440  
; Patent No. 6509455  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/09/657,440  
; CURRENT FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patent In. Ver. 2.0  
; SEQ ID NO 19:  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 8.4%; Score 43.6; DB 4; Length 38506;  
Best Local Similarity 51.0%; Pred. No. 1.6; Indels 0; Gaps 0;  
Matches 103; Conservative 0; Mismatches 99;  
QY 314 CAGCGTCAAGTCCCTGTCAGTCGATCGGCTCAGTCCGCGTCTGCCCTTGGCGCTGGG 373  
Db 37470 CAGGGGTGAGGCTGGGCGGAGGGGTGGCGCTCAGTCCGCGGCTCGCGCGGCGGCGG 37411  
QY 374 AGATAGCGGTTCACACAGCGGACACAGCGGCGGCGGCGGCGGCGGCGGCTTACGCCGA 433  
Db 37410 GGTGTGCGCGGCGGCGGCGGCTCAGCTCGGGAGCATCCCGCGGCGGCTGGCGGA 37351  
QY 434 TCGCTCGATGACACAGCGGCTCGGCGGCTCGGCGGCGGCTCGGCGATCCGTACCGCCCGGA 493  
Db 37350 AGTCTCTCTCGCGCAGCGGCTCGCGGCGGCTGGCGAGCGGTCTCGAGATCGCGA 37291  
QY 494 CCGCTCGGCGGCGGCGGCGG 515  
Db 37290 CGACGCGTCCCGCAGCGGCTG 37269

## RESULT 42

US-09-252-991A-4109  
; Sequence 4109, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4109:  
; LENGTH: 366  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4109

Query Match 8.3%; Score 43.4; DB 4; Length 366;  
Best Local Similarity 47.6%; Pred. No. 1.7;  
Matches 128; Conservative 0; Mismatches 141; Indels 0; Gaps 0;  
QY 11 GAAGGACCGGTGCTCGCGCGCGCGCATCACGAATGCCACTCCCGCGGGGTGACCAG 70  
Db 63 GATCGCACCTCTGTCGGCTGGAGCGGAGCCACTGTCGAGACATTTCCCTGGCCATCAT 122  
QY 71 GTCACACAGGTGGAGTTGAGCAGCAGCTCATCGCCCTGCGCCGCGCATGCCGAAACA 130  
Db 123 GGCCACCTCTGCTGCTGCTGTCGCGCGCGGAGTCAACGCTCTGGGTACAGCAGCGCATCGG 182  
QY 131 CCGGCGCAGGATCACGTGACAGCGGAGTGGCGCTCGATCTCGACGCTCGCGCCGAG 190  
Db 183 CAGCTCACTTCTCTTCGACCCCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 242  
QY 191 CTCGATCTGTCGCGCGCTCCACACAGGGGAACTGGCTCGGTGGCAGCGGAGCCG 250  
Db 243 CCGCGCATGCTCTTCGTCACCCCTGGCTGGCGAGGCGCTCGGCTACTTCAACAAC 302  
QY 251 CAGCGCGGACAGCTCGTCCAGGCTGGCGGT 279  
Db 303 CACCTGGCGCGGCTGCT 331

## RESULT 43

US-09-489-039A-123  
; Sequence 123, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 123:  
; LENGTH: 1287  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-123

Query Match 8.3%; Score 43.4; DB 4; Length 1287;  
Best Local Similarity 48.6%; Pred. No. 1.7;  
Matches 119; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
QY 40 TCAGGAATGCCACTCGCGCGGCTGACAGGTGCGACCGAGGTGGAGTTGAGCAGCAGC 99  
Db 672 TGACGATCGCGCGGAGGTTGGGCTGCTGCCAGATATTGTGGAACGTGCCAGCGACA 731  
QY 100 TCATCGCGGCTCGCGCGGATGCCGAAACACCGGCGCAGGATCAGTGCAGCAGCGCGA 159  
Db 732 GCTCATCCACAGGACACCGCGGAGTGGAGATAAAGCGCGAGCGCGCAGCG 791  
QY 160 TGGCGGCTCGATCTCGAGCGTGGCGGAGTGTCTGTCGCGCGGCTCCACACCA 219  
Db 792 CGAGCAGCGGTACGCGGATGTTCCCGTCCAGCGAGGCCACCAACCGCGCGCGCACCC 851  
QY 220 GGGGAACTGGCTCGGTGGCAGCGGAGCGCCCGGAGACCTCGTCCAGGCTGGCGT 279  
Db 852 AGCGCATTTGCCAGCGGAGCGCGCGCGGCTGGGACGCTTGGCCCTTCTGTTGATG 911  
QY 280 CACGG 284  
Db 912 AGCGG 916

## RESULT 44

US-08-740-223A-17/c

; Sequence 17, Application US/08740223A  
; Patent No. 6265564  
; GENERAL INFORMATION:  
; APPLICANT: Davis, et al.  
; TITLE OF INVENTION: Expressed Ligand - Vascular  
; TITLE OF INVENTION: Intercellular Signalling Molecule  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
; STREET: 777 Old Saw Mill Road  
; CITY: Tarrytown  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10591  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/740,223A  
; FILING DATE: 25-OCT-1996  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: USSN 60/022/999  
; FILING DATE: 02-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cobert, Robert J  
; REGISTRATION NUMBER: 36,108  
; REFERENCE/DOCKET NUMBER: REG 333  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 914-345-7400  
; TELEFAX: 914-345-7721  
; INFORMATION FOR SEQ ID NO: 17:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1512 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...1509  
; OTHER INFORMATION:  
; NAME/KEY: TIE ligand-4  
; LOCATION: 1...1512  
; OTHER INFORMATION:  
; US-08-740-223A-17

Query Match 8.3%; Score 43.4; DB 3; Length 1512;  
Best Local Similarity 49.8%; Pred. No. 1.7;  
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 44 GAATGCCACTCCGGCGGGTGACACAGGTGACAGGTGGAGTTGAGACGACGCTCAT 103  
Db 564 GAGCTTCTGCTCTGTAGCAGAGTGGTCTTCCAGCTTGTGTGGACAGAAAGGTCTC 505  
QY 104 CGCCGCTCGCGCGCATGCCAACAACCGGGCCAGGATCAGTGCAGCAGCGGATGCG 163  
Db 504 TGGCATCTGGGATCATCTTGTATGTTCTGGTTAGGAGCTGAGCTCCATGTCGGTCAG 445  
QY 164 CGCTCGATCTCGACGGTGGCGGCGAGCTCGATCTGTCGCCGGCTCCACACAGGGG 223  
Db 444 CTTGGGATCTGGGAGTGGTCTGTTTCAAGAGGCTGGTGGCCAGCTCTAGCATGGGGC 385  
QY 224 AAATGGCTCGGTGGCAGCGGCGAGCCCGGAGAGCT 264  
Db 384 CGTCTGATCTGGCCATTTGCTGTGACCTGCTCCAGCT 344

RESULT 45  
US-09-202-491-9/c  
; Sequence 9, Application US/09202491

; Patent No. 6432667  
; GENERAL INFORMATION:  
; APPLICANT: Valenzuela et al.  
; TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF  
; FILE REFERENCE: REG330-K  
; CURRENT APPLICATION NUMBER: US/09/202,491  
; CURRENT FILING DATE: 1998-11-16  
; EARLIER APPLICATION NUMBER: PCT/US97/10728  
; EARLIER FILING DATE: 1997-06-19  
; EARLIER APPLICATION NUMBER: 60/022,999  
; EARLIER FILING DATE: 1996-08-02  
; EARLIER APPLICATION NUMBER: 60/021,087  
; EARLIER FILING DATE: 1996-07-02  
; EARLIER APPLICATION NUMBER: 08/665,926  
; EARLIER FILING DATE: 1996-06-19  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1509)  
; US-09-202-491-9

Query Match 8.3%; Score 43.4; DB 4; Length 1512;  
Best Local Similarity 49.8%; Pred. No. 1.7;  
Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 44 GAATGCCACTCCGGCGGGTGACACAGGTGACAGGTGGAGTTGAGACGACGCTCAT 103  
Db 564 GAGCTTCTGCTCTGTAGCAGAGTGGTCTTCCAGCTTGTGTGGACAGAAAGGTCTC 505  
QY 104 CGCCGCTCGCGCGCATGCCAACAACCGGGCCAGGATCAGTGCAGCAGCGGATGCG 163  
Db 504 TGGCATCTGGGATCATCTTGTATGTTCTGGTTAGGAGCTGAGCTCCATGTCGGTCAG 445  
QY 164 CGCTCGATCTCGACGGTGGCGGCGAGCTCGATCTGTCGCCGGCTCCACACAGGGG 223  
Db 444 CTTGGGATCTGGGAGTGGTCTGTTTCAAGAGGCTGGTGGCCAGCTCTAGCATGGGGC 385  
QY 224 AAATGGCTCGGTGGCAGCGGCGAGCCCGGAGAGCT 264  
Db 384 CGTCTGATCTGGCCATTTGCTGTGACCTGCTCCAGCT 344

RESULT 46  
US-09-709-188-17/c  
; Sequence 17, Application US/09709188  
; Patent No. 6441137  
; GENERAL INFORMATION:  
; APPLICANT: Davis et al.  
; TITLE OF INVENTION: Expressed Ligand - Vascular Intercellular Signaling Molecule  
; FILE REFERENCE: REG 333-Z  
; CURRENT APPLICATION NUMBER: US/09/709,188  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 08/740,223  
; PRIOR FILING DATE: 1996-10-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 17  
; LENGTH: 1512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1512)  
; OTHER INFORMATION:  
; US-09-709-188-17

Query Match 8.3%; Score 43.4; DB 4; Length 1512;  
Best Local Similarity 49.8%; Pred. No. 1.7;



Matches 110; Conservative 0; Mismatches 111; Indels 0; Gaps 0;  
QY 44 GAACTGCCACTCCGGCGGGTGACAGGTCGACAGGTGGAGTTGAGCAGCCAGCTCAT 103  
Db |||||  
QY 564 GAGCTTCTGCTCTGAGCAGAGCTGGTTCTCCAGCTTGTGGTGACAGAAAGGTTCTC 505  
Db |||||  
QY 104 CGCGGCTCGCGCGGATGCGGAAACACCGGGCCAGGATCAGTGCAGCAGCGCGATCGG 163  
Db |||||  
QY 504 TGGCATCTGGGCATCCATTCTTGATGCTGTGGTTACGAGAGCTGAGCCTCCATGTCGTGTCAG 445  
Db |||||  
QY 164 CGGCTGATCTGACAGGTGGCGCGAGCTGATGCTGTCGCGGGTCCACACAGGGG 223  
Db |||||  
QY 444 CTTCGGGATCTGGGAGTGGTCTGTTTTCAGGAGGCTGGTGGCCAGCTCTAGCATGGGGC 385  
Db |||||  
QY 224 AAATCGCTCGGTGGCGAGCGGAGCGCCCGAGCGCGGACAGCT 264  
Db |||||  
QY 384 CGTCTGATCTGGGCCATTGCTGTGGACCTGCTCCAGCT 344  
Db |||||

RESULT 47  
US-09-489-039A-241/c  
; Sequence 241, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breston et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 241  
; LENGTH: 1773  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-241

Query Match 8.3%; Score 43.4; DB 4; Length 1773;  
Best Local Similarity 48.6%; Pred. No. 1.7;  
Matches 119; Conservative 0; Mismatches 126; Indels 0; Gaps 0;  
QY 40 TCACGAATCGCCACTCCGGCGGGTGACAGGTCGACAGGTGGGAGTTGAGCAGCCAGC 99  
Db |||||  
QY 1090 TGAGGATGCCCGCAGCAGGTTGGGCTGTGCGAGATATTGGAACTGCGCAGCGACA 1031  
Db |||||  
QY 100 TCTATCCCGCTGCGCGCGGATGCGGAAACACCGGGCCAGGATCAGTGCAGCAGCGGGA 159  
Db |||||  
QY 1030 GCTCATCCACAGCGCACGCGGAGCGCCCGAGTTGGAGATAAAGCGGCGAGCGCCAGC 971  
Db |||||  
QY 160 TGGCGCTCGATCTCGAGCTCGGCGGAGCTCGATCTCGTCCGCGCTCCACACCA 219  
Db |||||  
QY 970 CGAGCAGCGTACGCGGATGGTACGCTCAGCAGGCGACACCGCCCGCGCCAGCC 911  
Db |||||  
QY 220 GGGGAACTGGTTCGGTGCAGCGGAGCGCCCGAGCGGAGCAGCTCGTCCAGGCTGGCGT 279  
Db |||||  
QY 910 AGCGCATTTGCGCAGCGGAGCGCGCGCGCTGGAGAGCCCTTGCCTTCATGGTGACAA 851  
Db |||||  
QY 280 CACGG 284  
Db |||||  
QY 850 AGCGG 846  
Db |||||

RESULT 48  
US-08-460-269C-3  
; Sequence 3, Application US/08460269C  
; Patent No. 6197548  
; GENERAL INFORMATION:  
; APPLICANT: CLARE, JEFFREY J.  
; APPLICANT: ROMANOS, MICHAEL A.  
; TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST

NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Miller, White, Zelano & Branigan, P.C.  
STREET: 2000 Clarendon Blvd., Suite 1400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22201  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/460,269C  
FILING DATE: 02-Jun-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Lebovitz, Richard M.  
REGISTRATION NUMBER: 37,067  
REFERENCE/DOCKET NUMBER: Popov-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 243-6333  
TELEFAX: (703) 243-6410  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3000 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 145..2877  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-08-460-269C-3  
Query Match 8.3%; Score 43.4; DB 3; Length 3000;  
Best Local Similarity 45.4%; Pred. No. 1.7;  
Matches 193; Conservative 0; Mismatches 231; Indels 1; Gaps 1;  
QY 71 GTCACACAGGTGGGAGTTGAGCAGCCAGCTCATCGCGCTGCGCGGATGCGGCAACCA 130  
Db |||||  
QY 2182 GCCGACACACGGTGGCGGTGGCGGGCGGCTGGCACTGGCGGGCTGCGCGGTAT 2241  
Db |||||  
QY 131 CGGGCCAGGATCAGTGCAGCAGCGGATGCGGCTCGATCTCGACGTCGCGCGCAG 190  
Db |||||  
QY 2242 ACGGCGGCGACCGCGGCTTTACCGGCGACCGCGGCGGCGCACACCGACAG -CGTGCATGT 2300  
Db |||||  
QY 191 CTCGATCTCGTCCCGCTCCCAACAGGGGAACTGGCTCGGTGGCAGCGCGCAGCCC 250  
Db |||||  
QY 2301 CGGGGGCTATGCCACCTATATGCCAACAGCGTTTCTACTGACGCGACGTCGCGCG 2360  
Db |||||  
QY 251 CAGCGGAGCAGCTGTCAGGCTGGGTCACGGAATCTTCGCGGGTTCGGGAGCGGAAAC 310  
Db |||||  
QY 2361 CAGCGGCTCGAAATGACTTCAAGTGGCGGCGAGCGATGGGTAGCGGTTCAAGGCGAA 2420  
Db |||||  
QY 311 GCGCAGCGCTCAGATCCCTGTCAGTGCATCGGCTCAGTGCCTGCTCCCGCTTGGCCT 370  
Db |||||  
QY 2421 GTACCGACCCATGGGGTAGGCGCTTCGTCGAGGGGGCGCGGCTTCGCCCATGCCGA 2480  
Db |||||  
QY 371 GGAGAGATACGGTTTACAGCAGAGCGGCGACACCGCGGGCGGGCGGGCGGTTTCAGC 430  
Db |||||  
QY 2481 CGGCTGGTTCTCGAGCGCGCAGCGCGAGGTGGCGGTGTTCCGGGTGGCGCGGTTTCGTA 2540  
Db |||||  
QY 431 CGATCGCTCGATGACACAGCGGCTGGGGGTTCGGGCGGTCGCGGAGATCCGTACGCGCC 490  
Db |||||  
QY 2541 CCGGCGGCGCAATGGCTTCGGGTTCGCGAGCGGCGGCGAGTCGCTGCTGGGTGCCT 2600  
Db |||||  
QY 491 GGACC 495  
Db |||||  
QY 2601 GGGCC 2605  
Db |||||

## RESULT 49

US-09-344-510B-6/c  
; Sequence 6, Application US/09344510B  
; Patent No. 6579850  
; GENERAL INFORMATION:  
; APPLICANT: Nabeshima, Youichi  
; Kuroco, Makoto  
; Sekine, Susumu  
; Iida, Akihiro

TITLE OF INVENTION: No. 6579850el Polypeptide, No. 6579850el DNA and No. 6579850el  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10112-3801

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
COMPUTER: Compaq DeskPro EN  
OPERATING SYSTEM: Windows 98  
SOFTWARE: WordPad

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/344,510B  
FILING DATE: 25-Jun-1999

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/04585  
FILING DATE: 12-DEC-1997

APPLICATION NUMBER: JP 347871

FILING DATE: 26-DEC-1996

APPLICATION NUMBER: JP 205815

FILING DATE: 31-JUL-1997

ATTORNEY/AGENT INFORMATION:

NAME: Perry, Lawrence S.

REGISTRATION NUMBER: 31865

REFERENCE/DOCKET NUMBER: 766.32

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 218-2100

TELEFAX: (212) 218-2200

SEQUENCE CHARACTERISTICS:

LENGTH: 3163

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORGANISM: human

IMMEDIATE SOURCE:

LIBRARY: kidney

FEATURE:

NAME/KEY: CDS

LOCATION: 9..3047

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-344-510B-6

Query Match 8.3%; Score 43.4; DB 4; Length 3163;  
Best Local Similarity 53.9%; Pred. No. 1.7;  
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 189 AGCTCGATCTCGTCCGCGCTCCACACACAGGGGAACTGCTCGGTGGCGAGCGGCAGC 248  
DB 574 ACCACGGGCTGACGCCCGGCTCCGCGAGCGCTCCACAGCGCGCGGTAGTAGCGGAGC 515

QY 249 CCCAGCCGGGACAGCTCGTCCAGGCTGGCGTCAAGGACCTCTCGCGGTCCGGAGCGGAA 308  
DB 514 CCCTCGCGTTGGGACGCCCGCGCTGCCATTGGGGAGCACTCGCGGCCACGAGATGGAG 455

QY 309 AGCGGACGGCTCAGATCCCTGTGAGTCGATCGCATCGCTCAGTGGCG 353

DB 454 AAGCGGTAGTGAAGTACCCCGAGCTTCGCGACGCGCTCCGCTGCG 410

## RESULT 50

US-09-344-510B-7/c  
; Sequence 7, Application US/09344510B  
; Patent No. 6579850  
; GENERAL INFORMATION:  
; APPLICANT: Nabeshima, Youichi  
; Kuroco, Makoto  
; Sekine, Susumu  
; Iida, Akihiro

TITLE OF INVENTION: No. 6579850el Polypeptide, No. 6579850el DNA and No. 6579850el  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitzpatrick, Cella, Harper & Scinto  
STREET: 30 Rockefeller Plaza  
CITY: New York  
STATE: New York  
COUNTRY: United States  
ZIP: 10112-3801

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 1.44 mb, DS, DD  
COMPUTER: Compaq DeskPro EN  
OPERATING SYSTEM: Windows 98  
SOFTWARE: WordPad

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/344,510B  
FILING DATE: 25-Jun-1999

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP97/04585  
FILING DATE: 12-DEC-1997

APPLICATION NUMBER: JP 347871

FILING DATE: 26-DEC-1996

APPLICATION NUMBER: JP 205815

FILING DATE: 31-JUL-1997

ATTORNEY/AGENT INFORMATION:

NAME: Perry, Lawrence S.

REGISTRATION NUMBER: 31865

REFERENCE/DOCKET NUMBER: 766.32

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 218-2100

TELEFAX: (212) 218-2200

SEQUENCE CHARACTERISTICS:

LENGTH: 3435

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORGANISM: human

IMMEDIATE SOURCE:

LIBRARY: kidney

FEATURE:

NAME/KEY: CDS

LOCATION: 9..1655

IDENTIFICATION METHOD: E

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-344-510B-7

Query Match 8.3%; Score 43.4; DB 4; Length 3435;  
Best Local Similarity 53.9%; Pred. No. 1.7;  
Matches 89; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 189 AGCTCGATCTCGTCCGCGCTCCACACACAGGGGAACTGCTCGGTGGCGAGCGGCAGC 248  
DB 574 ACCACGGGCTGACGCCCGGCTCCGCGAGCGCTCCACAGCGCGCGGTAGTAGCGGAGC 515

QY 249 CCCAGCCGGGACAGCTCGTCCAGGCTGGCGTCAAGGACCTCTCGCGGTCCGGAGCGGAA 308  
DB 514 CCCTCGCGTTGGGACGCCCGCGCTGCCATTGGGGAGCACTCGCGGCCACGAGATGGAG 455

QY 309 AGCGGACGGCTCAGATCCCTGTGAGTCGATCGCATCGCTCAGTGGCG 353

Tue Jun 29 10:07:49 2004

db 454 AAGCGTATGTAGTACACCCGAGCTCGGCGACGCGCTCCGTCTCG 410

Search completed: June 27, 2004, 19:41:58  
Job time : 96.8824 secs

Result No.	Score	Query		ID	Description
		Match	Length		
1	520	100.0	109519	12	US-09-758-759-1
C 2	217	41.7	579	12	US-09-758-759-174
C 3	59.6	11.5	15738	15	US-10-329-079-46
C 4	59.6	11.5	61944	15	US-10-329-079-34
5	55.2	10.6	2514	13	US-10-329-079-34
6	54.8	10.5	3833	13	US-10-329-079-34
7	54.8	10.5	29833	13	US-10-329-079-34
8	52.4	10.2	88421	9	US-10-329-079-34
C 9	52.4	10.1	1059	16	US-10-329-079-34
C 10	51.6	9.9	1293	13	US-10-329-079-34
11	50.6	9.7	975	13	US-10-329-079-34
12	50.6	9.7	45055	13	US-10-329-079-34
C 13	50.4	9.7	661	16	US-10-329-079-34
C 14	49.6	9.5	1326	15	US-10-329-079-34

1 GGTGAGACGACGACCGGTGTCGCGCCGCGCGCCCATCAGAACTGCCACTCGCGG 60  
Qy  
109000 GGTGACGACGACGACCGGTGTCGCGCGCGCGCCCATCAGAACTGCCACTCGCGG 109059  
Db  
61 GGTTGACGAGTGCACGATGGAGATTGAGCAGCCAGTCTCATTCGCGCGCTCGCGCGCA 120  
Qy



Query Match 11.5%; Score 59.6; DB 15; Length 61944;  
Best Local Similarity 46.6%; Pred. No. 5.8e-06;  
Matches 191; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

US-10-329-079-34

1 GGTGACAGCAGAACACCGCGGTGGTGGCCCGCGCCCGCCATCAGCAACTGCCACTCCGCGG 60  
38215 GGTGTCGGGTAGAGAGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38156

61 GGGTACACAGGTGACACAGTGGAGTGGAGCAGCAGTATCATCCCGCCCTGCGCGGCA 120  
38155 CGGGCG 38096

121 TCCGAAACACCG 180  
38095 AGAGGGCTGATGCG 38036

181 TCG 240  
38035 TCG 37976

241 GCGGCG 300  
37975 CCG 37916

301 GAGCGGAAACCG 360  
37915 GCG 37856

361 CCTTGGCTGGAGAGTAGCGGTTCAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 410  
37855 CCGTGGCGGTCAACGGCAACCG 37806

## RESULT 5

US-10-282-122A-28178  
Sequence 28178, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangshu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELTRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 28178  
LENGTH: 2514  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
US-10-282-122A-28178

Query Match 10.6%; Score 55.2; DB 13; Length 2514;  
Best Local Similarity 47.5%; Pred. No. 0.00011;  
Matches 194; Conservative 0; Mismatches 213; Indels 1; Gaps 1;

109 CTTGCG 168  
242 CTTGCG 301

169 CGATCTCGAGCGGTGGCG 228  
302 CGATCAACAGATTTCTTCTGGCGAATACCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 361

229 GGTCTGCTGGCG 288  
362 GCG 421

289 CTGCGCGGTGGAGCGGGAACCG 348  
422 CCGGAGATCCG 481

349 TGGCGGTGCTGCTCCCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 408  
482 GGTCTCATCGGCAACGGCG 541

409 GCGGGCGCGCGCGCGGTTCAGCGGATCCGCTGCGATACCGCGGTGCGGGGTGCGGGCGCG 468  
542 GCG 600

469 GTGCGCGGAGATCCGTACCG 516  
601 CCGGCGGATTCGACGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 648

## RESULT 6

US-10-087-192-707  
Sequence 707, Application US/10087192  
Publication No. US20020182586A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
APPLICANT: Engelhard, Eric K.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
FILE REFERENCE: 529452000122  
CURRENT APPLICATION NUMBER: US/10/087,192  
CURRENT FILING DATE: 2002-03-01  
PRIOR APPLICATION NUMBER: US 09/747,377  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 2059  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 707  
LENGTH: 3833  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-087-192-707

Query Match 10.5%; Score 54.8; DB 13; Length 3833;

Best Local Similarity 46.2%; Pred. No. 0.00013;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCATCTCCGCGCGGTGACCAAGTGCACACAGTGGAGATTGAGCAGCCAGCTCATCG 105  
DB 758 AGTTCCGCGCGGCGCAAGTGAAGCAATCGTCGGGGTCTGTCGGCCACGACCCCATCAGCG 817  
QY 106 CGCGCTGCGCGCGCATGCGAAACACCGCGGCCAGGATCAGCTGACGACACGCGCATGCGCG 165  
DB 818 TCCCGCGCGCGCGCCACACCACTGCTGTAACCTGCCCCCAGCAGAGCGGCTG 877  
QY 166 CTTGATCTGACAGTTCGGCGCGAGCTGATCTGCTGCGCCCGGTCTCCACACAGGGGAA 225  
DB 878 TCGCGCGCTGCGGCAACGACAGCTGCGCGCCACACCCCGCGCGGCCAAAGTGCAGCTCGT 937  
QY 226 ACTGCTCGGTGGAGCGGCGACCCAGCGGACAGCTGCTCCAGGCTGCGCTCAGCGA 285  
DB 938 GCCGGTGGCACCGCCAGCGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 997  
QY 286 CTTTCGCGGTGCGAGCGGAAACGCGACAGCTGATCCCTGATCGATTCGATTCGATTCGCT 345  
DB 998 GCGCGAGCGCGGTGTCGGTGGTGGAGC-CGCCCTGAGCCCGCGCGCGCGCGCGCG 1056  
QY 346 CAGTCCCGCTGCTCCCTTGGCGTGGAGATAGCGTTTACGACGAGCGGACCGCG 405  
DB 1057 CTGAGCCGCTGCAACACCTGAGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 1116  
QY 406 CGCGCGCGCGCGCGCGGTTCAGCCGATCCGCTGATGACGCGGCTGCGGGTTCGGG 465  
DB 1117 CTGCGCGAGCGGCGCGCTGCAACACAGCGCTCCATGCTCCATGCGCGCGCGCG 1176  
QY 466 GCGGTGCGGAGATCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515  
DB 1177 CG 1226

RESULT 7  
US-10-087-192-706  
; Sequence 706, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 706  
; LENGTH: 23833  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(23833)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-087-192-706  
Query Match 10.5%; Score 54.8; DB 13; Length 23833;  
Best Local Similarity 46.2%; Pred. No. 0.0001;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCATCTCCGCGCGGTGACCAAGTGCACACAGTGGAGATTGAGCAGCCAGCTCATCG 105  
DB 10758 AGTTCCGCGCGGCGCAAGTGAAGCAATCGTCGGGGTCTGTCGGCCACGACCCCATCAGCG 10817  
QY 106 CGCGCTGCGCGCGCATGCGAAACACCGCGGCCAGGATCAGCTGACGACGCGCGATGCGCG 165

RESULT 8  
US-09-976-059-1  
; Sequence 1, Application US/09976059  
; Patent No. US20020164747A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
; FILE REFERENCE: 3019-PC1  
; CURRENT APPLICATION NUMBER: US/09/976,059  
; CURRENT FILING DATE: 2001-10-15  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 88421  
; TYPE: DNA  
; ORGANISM: Actinoplanes sp.  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (2077)..(3078)  
; OTHER INFORMATION: ORF 1; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (3118)..(4032)  
; OTHER INFORMATION: ORF 2; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (4038)..(5048)  
; OTHER INFORMATION: ORF 3; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (6665)..(5814)  
; OTHER INFORMATION: ORF 4; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (7703)..(6693)  
; OTHER INFORMATION: ORF 5; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9464)..(8130)  
; OTHER INFORMATION: ORF 6; negative strandedness  
; NAME/KEY: misc feature  
; LOCATION: (9691)..(10761)  
; OTHER INFORMATION: ORF 7; positive strandedness  
; NAME/KEY: misc feature  
; LOCATION: (12751)..(10829)

```
OTHER INFORMATION: ORF 8; negative strandedness
NAME/KEY: misc feature
LOCATION: (13617)..(12802)
OTHER INFORMATION: ORF 9; negative strandedness
NAME/KEY: misc feature
LOCATION: (15203)..(13614)
OTHER INFORMATION: ORF 10; negative strandedness
NAME/KEY: misc feature
LOCATION: (15531)..(15863)
OTHER INFORMATION: ORF 11; positive strandedness
NAME/KEY: misc feature
LOCATION: (15880)..(19035)
OTHER INFORMATION: ORF 12; positive strandedness
NAME/KEY: misc feature
LOCATION: (19032)..(39713)
OTHER INFORMATION: ORF 13; positive strandedness
NAME/KEY: misc feature
LOCATION: (39713)..(58900)
OTHER INFORMATION: ORF 14; positive strandedness
NAME/KEY: misc feature
LOCATION: (65826)..(66530)
OTHER INFORMATION: ORF 15; positive strandedness
NAME/KEY: misc feature
LOCATION: (66546)..(67370)
OTHER INFORMATION: ORF 16; positive strandedness
NAME/KEY: misc feature
LOCATION: (67384)..(70059)
OTHER INFORMATION: ORF 17; positive strandedness
NAME/KEY: misc feature
LOCATION: (70059)..(70662)
OTHER INFORMATION: ORF 18; positive strandedness
NAME/KEY: misc feature
LOCATION: (70659)..(71906)
OTHER INFORMATION: ORF 19; positive strandedness
NAME/KEY: misc feature
LOCATION: (73439)..(71964)
OTHER INFORMATION: ORF 20; negative strandedness
NAME/KEY: misc feature
LOCATION: (74216)..(73563)
OTHER INFORMATION: ORF 21; negative strandedness
NAME/KEY: misc feature
LOCATION: (75424)..(74213)
OTHER INFORMATION: ORF 22; negative strandedness
NAME/KEY: misc feature
LOCATION: (75535)..(76464)
OTHER INFORMATION: ORF 23; positive strandedness
NAME/KEY: misc feature
LOCATION: (78110)..(76449)
OTHER INFORMATION: ORF 24; negative strandedness
NAME/KEY: misc feature
LOCATION: (79824)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
NAME/KEY: misc feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
NAME/KEY: misc feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
NAME/KEY: misc feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
NAME/KEY: misc feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc feature
LOCATION: (85536)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
```

```
NAME/KEY: misc feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-758-059-1

Query Match 10.2%; Score 52.8; DB 9; Length 88421;
Best Local Similarity 46.5%; Pred. No. 0.00026;
Matches 240; Conservative 0; Mismatches 272; Indels 4; Gaps 2;

QY 7 GCACGAAACGACCGGTGTCGCCCGCCGCGCCATCAAGAACTGCGCACTCGCGCGGGTGA 66
DB 67518 GCACGAGACGTCGTGCGCGCCGCGCGCGATGAGCGCGCTGCGCGCGCGCGCGCA 67577
QY 67 CCAGGTCGACAGGTGCGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGCGCGCA 126
DB 67578 ACTCGCCCGCCGCTGTCGCCCGGAGCCCGTGGCGGGGGTGTCTCTCGCTACCGGA 67637
QY 127 AACACCGCGCCAGGATCACGTGTCAGCAGCGGCGATGCGCGCTCGATCTCGACGGTCCGCC 186
DB 67638 CGGCTCGCGCGACCTGCTGTCGCGCGCGCGCGCGCTGAGACCGCGCGCTGAT 67697
QY 187 GCAGCTCGATCTGTCGCCCGGCTCCACACAGGGGAACTGGCTCGGTGCGACGCGCA 246
DB 67698 CGCCCTGGCGCCCGCGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 67757
QY 247 GCGCCAGCGCGGACAGCTGTCAGGCTGGCGCTACGCGACCTCTTCGCGGGTTCGCGAGCGG 306
DB 67758 GCGCGCTCGCGCGCGCGCGCTGGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCG 67817
QY 307 AAACGGCGAGGCTCAGATCCCTGTGTCAGTCGATCGGCTCAGTGCCTGTCCTCCCTTG 366
DB 67818 CGAGCTGCGCGTGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 67875
QY 367 GCCTGGGAGGATAGCGGTTTCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 424
DB 67876 GCGAAGTCTCTCGCGCGGTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67935
QY 425 TTCAGCGGATCCGCTCGATGACAGCGGCTGCGGGGTTCGGGGCGGTTCGCGGATCCGTA 484
DB 67936 TTGGGCGCGCGCGCGCGCGCTCGGCTGTTGGCGGGGCGCTCCAGCGCGCGCTTCGCG 67995
QY 485 CCGCCCGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520
DB 67996 GCGCGCGCGGTCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 68031

RESULT 9
US-10-369-493-24313/c
Sequence 24313, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 24313
LENGTH: 1059
TYPE: DNA
ORGANISM: Deinococcus radiodurans
US-10-369-493-24313

Query Match 10.1%; Score 52.4; DB 16; Length 1059;
Best Local Similarity 48.6%; Pred. No. 0.00063;
Matches 143; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
```



```
QY 23 GTCCGCCGCGCCGCTACGAACTGCACTCGCGCGGGTGACCAAGTTCGACAGGTG 82
Db 795 GTCCGCCGCGGGTCCGCGAGGATCAGGCCGAGGACCAAGGATCAGCGCGCGAGGTC 736
QY 83 GGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCATGCCGAAACACCGCGCCAGGAT 142
Db 735 GGAAAGGATGTCGCGAAGAGTTCTCGGTGAGATCACTCGTAGCGGTGCGGTCGCG 676
QY 143 CAGTGTGAGCAGCGGATGCGGCTCGATCTGACGGTTCGCGCGCAGCTCGATCTCGTC 202
Db 675 CACGATCAGCATCGCCACGCTGTCCACGTAATCGTGTTCAGGTGAATCAGCGGTAGTC 616
QY 203 GCCCGGCTCCACACCAAGGGAACCTGGCTCGGTGCGAGCGGAGCCCAAGCCGCGACAG 262
Db 615 GCGGTACGCAAGCGCTGCGGTGCGCGCCACACTCGGACACTTCGAGCAGCTTCGCG 556
QY 263 CTCGTGACGCTGCGCTGACGAGCTTCGCGGGTCGGGAGCGGAACGCGGAC 316
Db 555 CTTGTCCACGCTGCTGACGCGGCTTCGCGTGTGCTCGCGCGCCGAGAGCGGAC 502

RESULT 10
US-10-329-027-17/C
; Sequence 17, Application US/10329027
; Publication No. US20030211567A1
; GENERAL INFORMATION:
; APPLICANT: STAFFA, Chris
; APPLICANT: ZAZOPOULOS, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS, METHODS AND SYSTEMS FOR DISCOVERY OF LIPOPEPTIDES
; FILE REFERENCE: 8822/2002
; CURRENT APPLICATION NUMBER: US/10/329, 027
; CURRENT FILING DATE: 2002-10-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 17
; LENGTH: 1293
; TYPE: DNA
; ORGANISM: Actinomyces
US-10-329-027-17

Query Match 9.9%; Score 51.6; DB 13; Length 1293;
Best Local Similarity 48.0%; Pred. No. 0.00096;
Matches 212; Conservative 0; Mismatches 224; Indels 6; Gaps 2;

QY 23 GTCCGCCGCGCCGCTACGAACTGCACTCGCGCGGGTGACCAAGTTCGACAGGTG 82
Db 522 GTCCGCCCTCGTCCAGCTCGCGAAGACCCGCTCGCGCGGGTCCGCGCGCGCGAGGC 463
QY 83 GGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCATGCCGAAACACCGCGCCAGGAT 142
Db 462 GCGTAGACTGCGCCACCGCGGGGCTGACGGAATGCGCGAAGCGTCCAGCAGGC 403
QY 143 CAGTGTGAGCAGCGGATGCGGCTCGATCTGACGGTTCGCGCGCAGCTCGATCTCGTC 202
Db 402 GTGTGTGAGCGGTGGAACACACAGCCGCGGCTCGCGCGCAGCAGCAGCGCTGGCG 343
QY 203 GCCCGGCTCCACACCAAGGGAACCTGGCTCGGTGCGAGCGGAGCCCAAGCCGCGACAG 262
Db 342 GCTACGCGGGCCCGGCCAGGCGCGAATCTCGGCTATCGCGCCATCGCGCCCTTCAGCGCCG 283
QY 263 CTCGTGACGCTGCGCTGACGAGCTTCGCGGGTCGGGAGCGGAACGCGGACAG --C 319
Db 282 GCGCGCGGCTTCGCGGTGCGGCTCGCGCGCAGGTCTCGACACTCCAGCGGCTCGGGTC 223
QY 320 TCAGATCCCTGTGATCGCATCGGCTGAGTCCGCGTCTCCCTTCGCGCTGGGAGATA 379
Db 222 GCGCGCGCGGACGAGCTCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 163
QY 380 CCGGTTACGACGAGCG ---GCACACGCGCGCGCGGGGCGGGGCGGTTTCAGCCGATCC 436
Db 162 CTCGCTCTCGCGCAGCGGTGCCGAGCCGCGCGGTTCGAGCAGCCCGGATCGCGCGCGCC 103
```

```
QY 437 GCTCGATGACCAAGCGCGGCTGCGG 458
Db 102 GATCTCGTAGTAGCAGCGCGAG 81

RESULT 11
US-10-107-431-14
; Sequence 14, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: STAFFA, Alfredo
; APPLICANT: ZAZOPOULOS, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-14

Query Match 9.7%; Score 50.6; DB 13; Length 975;
Best Local Similarity 50.9%; Pred. No. 0.0018;
Matches 145; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

QY 23 GTCCGCCGCGCGCCATCAGAACTGCACTCGCGCGGGTGACCAAGTTCGACAGGTG 82
Db 402 GCGCGCGCGCGCGACGACCACTAGCGCAAGGCCAAGCTCGCGGAGCCCGCGCGTGT 461
QY 83 GGAGTTGAGCAGCAGCTCATCGCGGCTCGCGCGCATGCCGAAACACCGCGGCGAGGAT 142
Db 462 GGAGGCCCGCGCGAGGGCGTCTGTCGAGCGGGCGTCTGCGGATCGCAATGTCGCGG 521
QY 143 CAGTGTGAGCAGCGGAGTTCG -CGCTCGATCTCGAGGTCGCGCGCAGCTCGATCTCGT 201
Db 522 CCGCGGACCCCGCGCGGTTCAGCTCTCGCGCGGTTCGCGCGGGGCTCGCGGAGCGGT 581
QY 202 CCGCGCGCTCCACACCAAGGGAACCTGGCTGGTGGCAGCGGCGAGCCCGCGGAGCA 261
Db 582 CCGCGCGGACACCTTCGCGCGGTTCGTCGAGCTGTCCAGCTGCGCGCCACCGGACTA 641
QY 262 GCTGTTCAGGTCGCGTCAAGGACCTTCGCGGGTCGGGAGCGG 306
Db 642 CGTCGACGTGCGCAGCTGTGGAAGCGGTGTCGCGCGCGCCACCG 686

RESULT 12
US-10-107-431-277
; Sequence 277, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: FARNET, Chris
; APPLICANT: STAFFA, Alfredo
; APPLICANT: ZAZOPOULOS, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 277
; LENGTH: 45055
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-277
```

Query Match: 9.7%; Score 50.6; DB 13; Length 45055;  
Best Local Similarity 50.9%; Pred. No. 0.001;  
Matches 145; Conservative 0; Mismatches 139; Indels 1; Gaps 1;  
QY 23 GTGCGCCCGCGCGCCCATACGAACTGCCACTCGCGCGGGGTGACCAAGTTCAGCAGGTG 82  
Db 44482 GCG 44541  
QY 83 GAGTTGAGCAGCAGCTATCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142  
Db 44542 GAGGCG 44601  
QY 143 CACGTGACGAGCG 201  
Db 44602 CCG 44661  
QY 202 CGCGCGCGCTCCACACAGCGGGAACCTGCTGCGTGGGAGCGCGCGCGCGCGCGCGCG 261  
Db 44662 CGCGCGCGGACACCTCCCGCGCGCTGCTGAGCGCGCGCGCGCGCGCGCGCGCGCG 44721  
QY 262 GCTGCTGTCAGGTGGCGGTACGAGCCTCTCGCGCGGTGCGGAGCGG 306  
Db 44722 CGTGGAGCTGCGCGAGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 44766

RESULT 13  
US-10-260-238-3954/c  
; Sequence 3954, Application US/10260238  
; Publication No. US20040016025A1  
; GENERAL INFORMATION:  
; APPLICANT: Budworth, Paul R.  
; APPLICANT: Moughamer, Todd G.  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiyaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION  
; FILE REFERENCE: 60111-NP  
; CURRENT APPLICATION NUMBER: US/10/260,238  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,448  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 6077  
; SEQ ID NO 3954  
; LENGTH: 661  
; TYPE: DNA  
; ORGANISM: *Triticum aestivum*  
US-10-260-238-3954

Query Match: 9.7%; Score 50.4; DB 16; Length 661;  
Best Local Similarity 47.5%; Pred. No. 0.0021;  
Matches 150; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 47 CTGCCACTCCGCGCGGGTACCAAGTTCAGACAGGTGGAGTTGACGACCGAGTTCATCGC 106  
Db 457 CCGGTCGTTCAGCGGCATCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 398  
QY 107 CGCTGCGCGCGCGATGCCGAACACCGGCGCAGGATCAGTGTGACGACGCGCGATCGCGC 166  
Db 397 CGCGGTGTCGCGGATGCTCAGTCTCCACACCGCGTCCGCGCGTCTGCCCATGAGCGC 338  
QY 167 CTCGATCTCGAGCTGCGCGCGCGAGTTCGATCTCTGTCGCGCGCGCTCCACCGGGGAA 226  
Db 337 CACCAGCGCTCGCGCGCGCTCGCTCTGCTCTCGAAGACCTCTCTCCGCGCACGCGCCTC 278

QY 227 CTGCTCGGTGGCAGCG 286  
Db 277 CCGCGCGGTGCGAGAAGCCCGCTTCAGGCTCGCAGGTCTCTCTCCACCATGCGGTGTC 218  
QY 287 CTCTCGGGGTGCGGAGCGGGAACCGCGCAGGCTCAGATCCCTGTCTAGTGCATTCGGCTC 346  
Db 217 CTCCCTCGTGAAGTCCGCTGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 158  
QY 347 AGTCCCGGTGCTCCCC 362  
Db 157 CGAGGCTTCATCACC 142

RESULT 14  
US-10-156-761-5142/c  
; Sequence 5142, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5142  
; LENGTH: 1326  
; TYPE: DNA  
; ORGANISM: *Streptomyces avermitilis*  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1326)  
US-10-156-761-5142

Query Match: 9.5%; Score 49.6; DB 15; Length 1326;  
Best Local Similarity 48.9%; Pred. No. 0.003;  
Matches 133; Conservative 0; Mismatches 139; Indels 0; Gaps 0;  
QY 43 CGAAGTCCCACTCGCGCGGGTACCAAGTTCAGACAGGTGGAGTTGAGCAGCGAGTCA 102  
Db 553 CGAGTACG 494  
QY 103 TCGCGCGCTCGCGCGCGCATGCCGAAACACCGCGCGCGAGGATCAGTGTGACGCGCGATGC 162  
Db 493 TCTCGATCAGCG 434  
QY 163 GCGCTCGATCTCGAGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 222  
Db 433 CCAGTGTGTTGAGGATGACAGGATCACGGTCTCCAGGAGCAGCGCATCCGCGAAGC 374  
QY 223 GAAACTGGCTTCGTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314  
Db 373 AGCTTCGACCG 314  
QY 283 GAGCTCTTCGCGGTTCG 314  
Db 313 AGATGTCG 282

RESULT 15  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          9.5%; Score 49.6; DB 15; Length 9025608;
Best Local Similarity 48.9%; Pred. No. 0.00082;
Matches 133; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 43 CGAAGTGCACCTCGCGGGGTGACAGAGTGCAGAGTGGAGTTGAGCAGCGCCAGCTCA 102
Db 6266553 CGAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266612

QY 103 TCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 162
Db 6266613 TCTCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266672

QY 163 GCGCCTCGATCTCGACCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 222
Db 6266673 CGAGTCTGCTGTGAGGATGACAGGATCAAGTCTCCAGGAGCAGCACTCCG 6266732

QY 223 GAAGTGCCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
Db 6266733 AGCCCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6266792

QY 283 GAGCCTCTCGCGGGTCCGCGAGCGGAAACGCGC 314
Db 6266793 AGATGTCCTCCCGCGAGCGGTACGAGCGGAC 6266824

RESULT 16
US-10-156-761-4136/c
; Sequence 4136, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          9.5%; Score 49.4; DB 15; Length 9025608;
Best Local Similarity 47.8%; Pred. No. 0.00092;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 2 GTGCAGCAGGAACGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61
Db 1215 GAGACGCTCGTAGAGTGGCGTGCCTCGCTCAGCCACAGGTGCGACCCCGCGGAT 1156

QY 62 GGTGACCAAGTCGACCAAGTGGGAGTTGAGCAGCGAGCTCATCGCGCGCTCGCGCG 121
Db 1155 GGCTAGCTGTGCGCGGTCCAGTGTGGTTCAGTCTGTCACCAAGTTGCGCTCGCG 1096

QY 122 GCCGAAACACCGCGGCAGGATCAGTGAGCAGCGCGGATGCGCGCTCGATCTCGAC 181
Db 1095 GACCCGCGCACCCAGCAGGTGCGCGCTCGGCACGAGGAGCAGCGACTCGGTCTC 1036

QY 182 CGCGCGCAGCTCGATCTCGTGCCTCGCGCTCCACACAGGAGGAAACTGCGTGG 241
Db 1035 CACCGCGAGTCTGCTGTCCTCCGACAGCAGCGCGCTAGCGGTGAACGCGGTAC 976

QY 242 CGCGAGCCCGACCGCGGAGCTCGTCCAGGTGCGGTGTCAGCGACCTCTCGCGGT 300
Db 975 CGCGCGTTCCAGCCAGGCGCGGTGCGCGGGGTGAGCGAAGCGGTACGCTCGGTCT 917

RESULT 17
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match          9.5%; Score 49.4; DB 15; Length 9025608;
Best Local Similarity 47.8%; Pred. No. 0.00092;
Matches 143; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 2 GTGCAGCAGGAACGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 61
Db 5099716 GAGACGCTCGTAGAGTGGCGTGCCTCGCTCAGCCACAGGTGCGACCCCGG 5099657
```

QY 62 GGTGACAGAGTGCACAGAGTGGAGTTGAGACAGCCAGCTCATCGCCGCTGCGCCGGCAT 121  
Db 5099556 GGTACAGCTGTCGCGGTCACAGTGGTGGTTCAGTCTGTCACAGAGTTGCGTTCGCGGCC 5099597  
QY 122 GCCGAAACACCGCGGCAGAGTACAGTGCAGCAGCGGATGCGGCTTCGATCTCGACGCT 181  
Db 5099596 GACCGGGACACCCAGCAGCTGCGCCCTCGGCACAGAGGACAGGACTGCGTCTCCAGCGC 5099537  
QY 182 CGCGCAGCTGATCTGTCGCGCGGCTCCACACAGGGGAACTGGCTCGGTGGCAG 241  
Db 5099536 CACCGCAGCTGCTGTCGCGGACACAGCGCTGACCGGTGACCGGTACCGGCGCAG 5099477  
QY 242 CGCGAGCCCGCAGCGGACAGCTGCTCCAGGCTGGGTCACAGGACCTCTCGCGGGTCGG 300  
Db 5099476 CCGCGCTTCAGCAGCGGCGAGGTGCGCGGGGTGACGGAACGCTACGCTCGGTGTCGG 5099418

## RESULT 18

US-10-282-122A-25414/c

; Sequence 25414, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; REMAINING Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

## Query Match

; Sequence 25414, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; REMAINING Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

## Query Match

; Sequence 25414, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELIUPA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIOR FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; REMAINING Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 25414

; LENGTH: 1575

; TYPE: DNA

; ORGANISM: Mycobacterium avium

US-10-282-122A-25414

## RESULT 20

US-10-369-493-39530/c

; Sequence 39530, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

Db 1542 GATGTCCAGCACACGCGGTTGACCTCGGGACCTCGTGTGATGCGGGTCGAGATCCG 1483  
QY 104 CGCGCTGCGCGCGCATGCCGAAACACCGCGGCCAGGATCACGTGCAGCAGCGGATGCG 163  
Db 1482 CTCAGCAGCTCTGAGGAGCCCGGTCAGTCCGCGGTCTATGCGTCTCTACTGGTCAC 1423  
QY 164 CGCTCGATCTGAGGTCGCGCCAGCTCGATCTGTCGCGCGGTCTCCACACAGGGG 223  
Db 1422 CGCGCGCAGCAGTACGATGCGGTCGCGGTCTGTCGCGCTGCGCGCGCGCGCG 1363  
QY 224 AAATGGCTCGGTGCGCAGCGGACCGCCAGCGCGGACAGCTCGCTCCAGGTCGCGTCA 283  
Db 1362 GACCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1303  
QY 284 GACCTC 289  
Db 1302 GAGTTC 1297

## RESULT 19

US-10-369-493-39160/c

; Sequence 39160, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 39160

; LENGTH: 705

; TYPE: DNA

; ORGANISM: Xanthomonas campestris

US-10-369-493-39160

## Query Match

; Score 9.4%; DB 16; Length 705;

; Best Local Similarity 52.2%; Pred. No. 0.0046;

; Matches 109; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 154 CGCGATGCGCGCTCGATCTCGACGCTGCGCGCGCAGCTCGATCTCGTCCGCGCTCCC 213

Db 397 CGCGCGCGCTCCACAGTTCGTTGAGCAGCCGCGCGCGCGCGCGCGCGCGCGCG 338

QY 214 ACACAGCGGAACTGCTCGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 273

Db 337 ACACAGCGGAACTGCTCGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278

QY 274 TGGGTGTCAGGACCTCTCGCGGTCGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 333

Db 277 CGGTGGTCCCGAGACAGCGGTTTCCGCGCTGACGCAACAGCGGATCGCGGCA 218

QY 334 GTCCATCGGCTCAGTCCGCTCGTCCCG 362

Db 217 AGCCCTGCGCGAGCTGAAGTCGATGCC 189

## RESULT 20

US-10-369-493-39530/c

; Sequence 39530, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 39530  
; LENGTH: 711  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-39530

Query Match 9.4%; Score 49; DB 16; Length 711;  
Best Local Similarity 52.2%; Pred. No. 0.0046;  
Matches 109; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 154 CGGGATCGGGCTCGATCTCGAGCTGGCGCGACGCTGATCTGTCGCCCGGCTCCC 213  
DB 406 CGGGCGGGCGCTCCACCACTTCGGTGAGACCGCGCCAGCGCTGAAATTGGACGCC 347  
QY 214 ACACAGGGGAACTGGCTCGGTGGCAGCGGACGCGCCCGCGGACAGCTCGTCCAGGC 273  
DB 346 ACACAGCGGAATCTGGCCGCTGCTTGACGACGCGCGCGCTGGCGCTCGTCCAGCC 287  
QY 274 TGGCGTCACGACCTCTCGCGGGTGGGAGCGGAAACGCGACGCGCTCAGATCCCTGTCA 333  
DB 286 CGGTGGTGGCGGACGACCGAGCTTGGCGGCTGCACGACGACGCGGATCGGGCAA 227  
QY 334 GTCCGATCGGCTCAGTGGCGGTCTGCTCCC 362  
DB 226 AGCCTGGCGGACGCTGAAGTCGATCGCC 198

RESULT 21  
US-10-369-493-39909/c  
; Sequence 39909, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 39909  
; LENGTH: 711  
; TYPE: DNA  
; ORGANISM: Xanthomonas campestris  
US-10-369-493-39909

Query Match 9.4%; Score 49; DB 16; Length 711;  
Best Local Similarity 52.2%; Pred. No. 0.0046;  
Matches 109; Conservative 0; Mismatches 100; Indels 0; Gaps 0;  
QY 154 CGGGATCGGGCTCGATCTCGAGCTGGCGCGACGCTGATCTGTCGCCCGGCTCCC 213  
DB 406 CGGGCGGGCTCCACCACTTCGGTGAGACCGCGCCAGCGCTGAAATTGGACGCC 347  
QY 214 ACACAGGGGAACTGGCTCGGTGGCAGCGGACGCGCCCGCGGACAGCTCGTCCAGGC 273  
DB 346 ACACAGCGGAATCTGGCCGCTGCTTGACGACGCGCGCGCTGGCGCTCGTCCAGCC 287  
QY 274 TGGCGTCACGACCTCTCGCGGTGGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTCA 333

DB 286 CGGTGGTCCCGAGACCGAGCGTTTCCCGCTGCACGCAACAGCCAGGATCGGCGAA 227  
QY 334 GTCCGATCGGCTCAGTGGCGGTCTGCTCCC 362  
DB 226 AGCCTGGCGGAGGCTGAAGTCGATCGCC 198

RESULT 22  
US-09-758-759-128/c  
; Sequence 128, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Everniminacin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 128  
; LENGTH: 858  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(858)  
; OTHER INFORMATION: evbt  
US-09-758-759-128

Query Match 9.4%; Score 49; DB 12; Length 858;  
Best Local Similarity 51.1%; Pred. No. 0.0045;  
Matches 115; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 29 CGGGCGGGCCATCAGCAACTGCCTCCGCGGGGTGACAGGTGACAGGTGGAGTT 88  
DB 720 CCCCTGAGGTAGACGGTGGCGGCTCGGTGGGTCGCCGCGGAGAGCGGAGTC 661  
QY 89 GAGCAGCGAGTCTATCGCGGCTGGCGCGGATGCCGGAATCCGAAACACCGGGCCAGGATCACGTG 148  
DB 660 GAGCAGCAGCAAGTGTCCGCGCTCGCGCGCCAGCGCGCGCGCGCGCGAGCAC 601  
QY 149 CAGCAGCGGATGCGCGCTCGATCTCGACGTGCGCGCAGCTCGATCTGTCGCCCGG 208  
DB 600 CCGTACGGTGTGTTGTCGGCTTCGCCAGCGGACGAGGTGCCGAGCTGTCGGCCAT 541  
QY 209 CTCCACACACAGGGAAACTGCTCGGTGGCAGCGCGAGCCCCAG 253  
DB 540 CACCGATCGGCGCGCACCGGCGCGCGCAGCACCGCCTCGCTGAG 496

RESULT 23  
US-10-152-886-14/c  
; Sequence 14, Application US/10152886  
; Publication No. US20030064491A1  
; GENERAL INFORMATION:  
; APPLICANT: ECOPIA BIOSCIENCES INC.  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staifa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING  
; FILE REFERENCE: 3011-3US  
; CURRENT APPLICATION NUMBER: US/10/152,886  
; CURRENT FILING DATE: 2002-05-21  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 5760

```

Qy 131 CCGGGCCAGGATCAGTGCAGCAGCGGCGATCGCGCGCTCGATCTCGACGGTTCGGCGCGAG 190
Db 12426 CAGGCGGTCCCGGGGAAAGACGTGCGCGAAGTGCAGCGACGGGTTCGACCTTTGGCGCGAG 12367
Qy 191 CTCGATCTCGTCCGCGGGCTCCCAACACAGAGGGAACCTGGCTCGGTGGCAGCGGCAGCCC 250
Db 12366 GGTGGCTTCAGTTTCGCGCTCGGTCTATTCGGGCGAGCGCGGTGACTGAACACCATCGC 12307
Qy 251 CAGCGGGGACAGCTCGTCCAGCGCTGGGTTCAGCGACCTTCGCGGGTCGGGAGCGGAAAC 310
Db 12306 CAGGTGCACACGCGGTGATCGCGCCGTAGGCGGCGGACGACCTCGTCCCGCATCGCGGC 12247
Qy 311 GCGCAGGC 319
Db 12246 GAGCGGTC 12238

RESULT 25
US-10-132-134-1/c
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Parnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 52101
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. rosaceus
US-10-132-134-1

Query Match 9.4%; Score 49; DB 15; Length 52101;
Best Local Similarity 49.8%; Pred. No. 0.0024;
Matches 124; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy 71 GTCGACACAGTGGAGTTGAGCAGCAGCTCATCGCCGCTCGCGCGCATGCCGAAACA 130
Db 30119 GTCCGCTGTGGAGTTCCGATGAAGCTCACGAGGAGGATCAGCAGCAGAGTC 30060
Qy 131 CCGGGCCAGGATCAGTGCAGCAGCGCGATGCGCCCTCGATCTCGACGGTTCGGCGCGAG 190
Db 30059 CAGGCGGTCCCGGGGAAAGACGTGCGGAAAGTGCAGCGGTCGACCTTTGGCGCGAG 30000
Qy 191 CTCGATCTCGTCCCGCGGTCCCAACACAGAGGGAACCTGGCTCGGTGGCAGCGGCAGCCC 250
Db 29999 GGTGGCTTCAGTTTCGCGCTCGGTATCCGGGCGAGCGCGGTGACTGAACACCATCGC 29940
Qy 251 CAGCGGGGACAGCTCGTCCAGCGTGGCGGTTCAGCGACCTTCGCGGGTCGGGAGCGGAAAC 310
Db 29939 CAGGTGCACACGCGGTGATCGGCGGTAGGCGGCGGACGACCTCGTCCCGCATCGCGGC 29880
Qy 311 GCGCAGGC 319
Db 29879 GAGCGGTC 29871

RESULT 26

```

```

US-09-758-759-1/C
; Sequence 1, Application US/09758759
; Publication NO. US20040101832A1
; GENERAL INFORMATION:
; APPLICANT: Hosted, Thomas J.
; APPLICANT: Wang, Tim X.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Evernimomicin Biosynthetic Genes

```

FILE REFERENCE: ID0983K US  
CURRENT APPLICATION NUMBER: US/09/758,759  
CURRENT FILING DATE: 2001-01-11  
PRIOR APPLICATION NUMBER: US 60/175,751  
PRIOR FILING DATE: 2000-01-12  
NUMBER OF SEQ ID NOS: 204  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 1  
LENGTH: 109519  
TYPE: DNA  
ORGANISM: Micromonospora carbonacea  
US-09-758-759-1

Query Match 9.4%; Score 49; DB 12; Length 109519;  
Best Local Similarity 51.1%; Pred. No. 0.0022;  
Matches 115; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 29 CGGCGCGCCATCAGAACTGCCTCGCGCGGGTGACAGGTCGACCGAGTGGAGTT 88  
DB 82792 CCCCTCGAGCTAGAGCGTGGCGGCTCGGTGCGGTCGCGAGCGCGGAGCGGAAATC 82733

QY 89 GAGCAGCGAGTCTATCGCGGCTGCGCGGATGCGGAAACACCGCGCCAGGATCACGTG 148  
DB 82732 GAGCAGCAGAACTGCGCGGCTGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCAG 82673

QY 149 CAGCAGCGGATCGCGGCTCGATCTCAGCGTGGCGCGGCTCGATCTCGTGGCGGG 208  
DB 82672 CGGTACGGTGATGTTGTCGGCTTCGCCAGCGCGCAGGTTGCGGAGCTGTCGGCCAT 82613

QY 209 TCCCCACACACAGGGGAACTGCTCGTGGCGAGCGCGGCGCGCCAG 253  
DB 82612 CACCGATGCGCGCGCGCACCGGCGCGCGCGCGCTCGCTGAG 82568

RESULT 27  
US-10-156-761-6291/c  
Sequence 6291, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 6291  
LENGTH: 1116  
TYPE: DNA  
ORGANISM: Streptomyces avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(1116)  
US-10-156-761-6291

Query Match 9.4%; Score 48.8; DB 15; Length 1116;  
Best Local Similarity 49.4%; Pred. No. 0.0048;  
Matches 159; Conservative 0; Mismatches 157; Indels 6; Gaps 1;

QY 39 ATCAGAACTGCCTCGCGCGGTTGACCGAGTGCACCGAGTGGAGTTCAGCGCCAG 98  
DB 524 ACCATCGGTAGCGTTCGAGCGAGATCAGCGTTCAGCTGATCGCGAGCGG 465  
QY 99 CTCATCGCGCGCTGCGCGGATGCCGAAACACCGCGGCGAGATCACGTGACGCGCG 158

DB 464 ATGACCTCGTGATCGCGGGATGCCGCTTGGTGGCGGGATCTTGATGAGCGTGTG 405  
QY 159 ATGCGCGCTCGATCTCGACGCTCGCGCGCAGCTCGATCTCGTCCGCGGCTCCACACC 218  
DB 404 GGGCGGTGACCGAGCAG-----GCCAGCTCTTGGGCTCGGCGACCGTCCGCTTGGTG 351  
QY 219 AGGGAAACTGGCTGGTGGCAGCGGCGAGCCGCCAGCGGAGCAGCTCGTCCAGGCTGGCG 278  
DB 350 TTGTGGCCAGCGCGGTCGACCTCGATCGAGACCCCGCGCTCTGGCCGTCGCTGGCG 291  
QY 279 TCAGGACCTCTCGCGGTCGAGGAGGAGGAAACGCGCAGCGCTCAGATCCCTTCAGTGGC 338  
DB 290 TCGAAGACCGCGCGAGGATGTCGCGCGGCGTCCGCGAGCTCCGCGCTGATCATGCGG 231  
QY 339 ATCGGCTCAGTCCCGTCTGCTCC 360  
DB 230 ATGGCTTCTTCGACGGTGACCC 209

RESULT 28  
US-10-282-122A-14525  
Sequence 14525, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: Patent In version 3.1  
SEQ ID NO 14525  
LENGTH: 2208  
TYPE: DNA  
ORGANISM: Burkholderia mallei  
US-10-282-122A-14525

Query Match 9.4%; Score 48.8; DB 13; Length 2208;  
Best Local Similarity 49.6%; Pred. No. 0.0044;  
Matches 125; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

65	Qy	GACCAGTGCACCAAGTGGGAGTTGAGCAGCCAGCTCATGCCCGCTGGCGCGGCGATGCC	124
826	Db	GATGGGGGACACGTCGAGCACTTCTTGGCTGGCGGACACCGTGTCTGCCGAATCCG	885
125	Qy	GAACAACCGGGCCAGGATCAGTCTCAGCA CGCGATGGCGGCTCGATCTCGACGGTCGG	184
886	Db	AACGGCACGATCTCGGACGACCTGTACACGGCCGATCGAACTTCGACCAATTACCGAAG	945
185	Qy	CCGAGCTCGATCTCTGCTGGCCCGCTCCCA CACAGGGGAACTGGCTCGTGTGCGACGG	244
946	Db	AAGCAGTGTGCTGTCCGCTATCATGTTCTGAGCACAACTGAACCCGCTGTGACGCTCGG	1005
245	Qy	CACCCCAGCCGGACAGCTCTGTCCAGCTGGCGCTCAGGACCTCTCGCGGTCGGGAGC	304
1006	Db	CAGAACGTGGCTGGATGCACCTGTCTCTCGACGACGCTCGTCTACGGCGGCGGCTC	1065
305	Qy	GGAAACGCGCAC	316
1066	Db	GACGACGACAGC	1077

```

RESULT 29
US-10-156-761-7305/c
; Sequence 7305, Application US/10156761
; Publication No.: US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7305
; LENGTH: 654
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(654)
US-10-156-761-7305

```

Db 402 GCGCGCGCGCACTTGTGGACGTCGCCGCC 372

RESULT 30

US-10-282-122A-28770/c

; Sequence 28770, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

	Query Match	9.3%;	Score 48.6;	DB 13;	Length 1797;
	Best Local Similarity	47.7%;	Pred. No. 0.005;		
	Matches 173;	Conservative	0;	Mismatches 189;	Indels 1;
				Gaps	1;
QY	154	CGSCGATGCGGCTCGATCTCGACGTCGGCGCAGTCGATCTCGTCCGCCGGCTCCC	213		
DB	1768	CGCCGTGTCTGCCCCGATCAACGAGTTCTTCTGCGAATACCGGGCGCCGCTGATCG	1709		
QY	214	ACACCAAGGGAACCTGGCTCGGTGGACGCGCAGCCCCAGCCGGGACAGCTCTCCAGGC	273		
DB	1708	GCAACGGCACCAACGCGGCCCGGCACCGGGGCCAATGCGCGGGACGGCGGTGGTTAA	1649		
QY	274	TGSCGTACGCGACTCTCGCGGTCGGGACGGGAACGCGCAGCGCTCAGATCCCTGTCA	333		
DB	1648	TGGGCAACGCGCGCGCGAGAGATCCGCGCGCGCGCGGTCAACGGCGGGCGCGCGCGCA	1589		
QY	334	GTGCGATCGGCTCAGTGC CGGTCGTCCCTTGGCTGGGAGGATACGCGTTTCACACGA	393		
DB	1588	ACGGCGGCGCGCGCGGGCTCATCTGGCAACCGGCGGCGCGCGCGCGCGGAGGGCCA	1529		





Qy	123	CCGAAACACCGGGCCAGATCACTGTCAGACACGGCCGATGCGCGCTCGATCTTCGACGGTC	182
Db	3479	CTGCGCGCCGGATACATCCGCTTCAGCGCCACCTCGGCCGAACTCGGSCAGCGTGATCGGC	3420
Qy	183	GGCGCAGCTCGATCTCGTCCCGCGCTCCACACACGAGGGAACTGGCTCGGTGGCAGC	242
Db	3419	GAGAGCCGCGAGTGGCGCGCGACCGGCGCCGACACCTCGTGATGCGCGCGCGCGCGCAC	3360
Qy	243	GGCAGCCCCCAGCCCGGACAGCTCTCCACAGGTGCGCTCACGCACTCTCCGCGGTCGGGA	302
Db	3359	AGCAGCGCGCAGCGGGCGACCGCCACAGCGCGCAGGCGCGTTCGGGACAGCGCCCGTAG	3300
Qy	303	GCGGAAACGCGCACGGCTCAGATCCCTGTGTCAGTCGCATCGGCTCAGTCCGGTCTGTCCC	362
Db	3299	CGGTGACGACGATTCCTCGACACCGC-GTCGATCTCGTCTCGAGCCCGCGCGGCCAG	3241
Qy	363	C-----TTGGCCTGGGAGGATACGGGTTCACGACGACGGCACCGCGCGGGGGC	416
Db	3240	CCGCCGATACGCCCTCCACCGCACCGCATCGCTGGCGATAGTCCGCGCGCAGGTGCGC	3181
Qy	417	GGGGCGCGTTTCAGCCGATCCGCTCGATGACACAGCGGCTCGCGGGTCCGGGCGGTTCGGCGA	476
Db	3180	GTCCACCGCAGGTTCGATCCGCACGTCTCTTGGGTCTCTCGCGGTTGTGACGCTCTGGCC	3121
Qy	477	GATCCGTACCGCCCGGACCGGCTCGGCCACGCGCGCGGATCC	520
Db	3120	GTCCGCGCGCGCCCGGTACGCTTCGACGGGCTCGCCACACAGCC	3077

RESULT 33

```

US-10-369-493-41689/c
; Sequence 41689, Application US/10369493
; Publication NO. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 39-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 41689
; LENGTH: 1350
; TYPE: DNA
; ORGANISM: SPHINGOMONAS
US-10-369-493-41689

```

Query Match	9.3%	Score 48.2	DB 16	Length 1350
Best Local Similarity	47.2%	Pred. No. 0.0066		
Matches 146	Conservative 0	Mismatches 163	Indels 0	Gaps 0
Qy	64	TGACCAAGTCGACCAAGGTGGAGTTGAGCAGCAGCAAGTCATCGCGCGCTGCGCGCGGATGC	123	
Db	1108	TGGCCAGCTCGGCGATCTCTGAGCCGCGAGGACCCCGCCAGCTTGCACGAAAGTC	1049	
Qy	124	CGAAACACCGGGCCAGGATCAGTTCGACGACCGCGATGCGCGGCTCGATTCGACGGTCG	183	
Db	1048	CGGTAAATGCCGGTCAACAGTACCGCGCGCGCGTGGCGCGCGAGACCTTGGCAGCGCGC	989	
Qy	184	GCGCAGCTCGATCTCGTTCGCGCGCGTCCACACACAGGGGAAACTGCTCGGTGGCAGCG	243	
Db	988	GGGGACCAAGCGCTCAGCGGCCATCAGAAAGATGCGGCTCTGGCGTACATCATGA	929	
Qy	244	GCAGCCCCACCGGGAAGTCTGCTCAGGTGCGCTACGGAACCTTCGCGGTGCGGAG	303	
Db	928	CCAGGATCACCGAGGGCAGCGCACCAAGCGCGCGCGGCGATCGCCAGCGCGCGCAAGG	869	

QY 304 CGGAACGCGCAGGGCTCAGATCCCTGTCACTGCATCGGCTCAGTGCCTGTCGTTCCCC 363

Db 868 GGTGGCGGAGCGTGCGCAGCACCGGCCAGCGGCTCGGGCGAATTGCCGAGCGGTGCGCG 809

QY 364 TTGSCCTGG 372

Db 808 TCGCATCG 800

RESULT 34  
US-10-338-110-119  
; Sequence 119, Application US/10338110  
; Publication No US20040023254A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuhrmann, Jeffrey J.  
; APPLICANT: Romesser, James A.  
; TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial  
; TITLE OF INVENTION: Communities  
; FILE REFERENCE: HER-0056  
; CURRENT APPLICATION NUMBER: US/10/338,110  
; CURRENT FILING DATE: 2003-01-07  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 119  
; LENGTH: 536  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus Sequence  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (7)..(7)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (21)..(21)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (69)..(69)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (87)..(87)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (213)..(213)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (226)..(226)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (261)..(261)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (273)..(273)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (347)..(347)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (357)..(357)



```
QY 143 CAGCTGACAGCGGCGATCGGCGCTCGATCTCGACGGTGGCGCGGAGTCGATCTCGTC 202
Db 2913 CCGCGCCAGGTCTCTCCCGCGGAGCGCGAGTAGAGAT-----CACACCGCACGTC 2860
QY 203 GCCCGGCTCCACACACAGGAGAACTGGCTCGGTGCGACGGCAGCCCGCAGCGGAGACAG 262
Db 2859 CTCGGGTCGGGACCTTGGCGATCCAGCGCCCGCGCGCGCGCGCGCGCGCGCGCG 2800
QY 263 CTGCTCAGAGCTGGGCTCAGGACCTCTCGCGGTCGGGAGGAGAAACCGCGACGGCTCA 322
Db 2799 CAGCGGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2740
QY 323 GATCCCTGTGTCAGTCGATCGGCTCAGTCGCGGTCTGCTCCCTTGGCTCGGAGGATAG 382
Db 2739 CGAGGCGAGCGCGCGGTCGACATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 2680
QY 383 GTTCACGACAGCGGCGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
Db 2679 GTGCTCGGTGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2620
QY 443 TGACCA-----CGGCTCGGCGGTGCGGCGCGGTGCGGCGGATCGGTACCGCGCG 496
Db 2619 GGGCGCGGCGGTGCGTGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2560
QY 497 CCGCGCGCGCGCGCGCGCG 516
Db 2559 GCGCTCGCGCGCGCGCGCG 2540

RESULT 37
US-09-827-688-8
; Sequence 8, Application US/09827688
; Publication No. US20030165476A1
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 9.2%; Score 48; DB 10; Length 154746;
Best Local Similarity 46.6%; Pred. No. 0.0037;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
Db 129000 GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129059
QY 83 GGAATTGAGCAGCAGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
Db 129060 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129119
QY 143 CAGCTGACAGCGGCGATCGGCTCGGTCGAGTCTCGACGGTGGCGCGCGCGCGCGCG 202
Db 129120 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129173
QY 203 GCCCGGCTCCACACACAGGAGAACTGGCTCGGTGCGACGGCAGCCCGCAGCGGAGAC 262
Db 129174 CTCGGGTCGGGACCTTGGCGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 129233
QY 263 CTGCTCAGAGCTGGGCTCAGGACCTCTCGCGGTGCGGAGCGGAGCGGAGCGGAG 322
```

```
Db 129234 CAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129293
QY 323 GATCCCTGTGTCAGTCGATCGGCTCAGTCGCGGTCTGCTCCCTTGGCTCGGAGGATAG 382
Db 129294 CGAGGCGAGCGCGCGCGGTGCGACATGAGGCGCGCGCGCGCGCGCGCGCGCG 129353
QY 383 GTTCACGACAGCGGCGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
Db 129354 GTGCTCGGTGAGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129413
QY 443 TGACCA-----CGGCTCGGCGGTGCGGCGCGGTGCGGCGCGCGCGCGCGCGCG 496
Db 129414 GGGCGCGCGCGGTGCGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129473
QY 497 CCGCGCGCGCGCGCGCGCG 516
Db 129474 GCGCTCGCGCGCGCGCGCG 129493

RESULT 38
US-09-827-688-8/c
; Sequence 8, Application US/09827688
; Publication No. US20030165476A1
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERMA
; APPLICANT: BHOGAL, BALBIR
; TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; FILE REFERENCE: P01949US1/10004014
; CURRENT APPLICATION NUMBER: US/09/827,688
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 9.2%; Score 48; DB 10; Length 154746;
Best Local Similarity 46.6%; Pred. No. 0.0037;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;

QY 23 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 82
Db 152742 GTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152683
QY 83 GGAATTGAGCAGCAGCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 142
Db 152682 GCAGGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152623
QY 143 CAGCTGACAGCGGCGATCGGCTCGGTCGAGTCTCGACGGTGGCGCGCGCGCGCG 202
Db 152622 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152569
QY 203 GCCCGGCTCCACACACAGGAGAACTGGCTCGGTGCGACGGCAGCCCGCAGCGGAG 262
Db 152568 CTCGGGTCGGGACCTTGGCGATCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 152509
QY 263 CTGCTCAGAGCTGGGCTCAGGACCTCTCGGCGGTGCGGAGCGGAGCGGAGCGGAG 322
Db 152508 CAGCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152449
QY 323 GATCCCTGTGTCAGTCGATCGGCTCAGTCGCGGTCTGCTCCCTTGGCTCGGAGGATAG 382
Db 152448 CGAGGCGAGCGCGCGCGGTGCGAATGAGGCGCGCGCGCGCGCGCGCGCGCGCG 152389
QY 383 GTTCACGACAGCGGCGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
```

Db 152388 GTGTCGTGAGCTCGGCACCGCCCGGGGGCGCAGTAGGCTCCAGGGCGGCGGCGGA 152329  
Qy 443 TGACCA-----GGCGGTTCGGGGTTCGGGGCGGTTCGGGAGATCCGTACCGCCCGACCG 496  
Db 152328 GGGCGCGCGGTGTGTGGCTGGGCCCGCGGTGGCGCGCCAGCGCCCTTCGCGGTTCGGG 152269  
Qy 497 CCTCGGCAGCGCGCGCGG 516  
Db 152268 GCCCTCGGCGCGCGCGG 152249  
RESULT 39  
US-10-156-761-2330/c  
; Sequence 2330, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2330  
; LENGTH: 1341  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1341)  
US-10-156-761-2330  
Query Match 9.2%; Score 47.8; DB 15; Length 1341;  
Best Local Similarity 46.0%; Pred. No. 0.0083;  
Matches 199; Conservative 0; Mismatches 232; Indels 2; Gaps 1;  
Qy 76 CAGGTGGAGTTGAGCAGCAGCTATCGCCGCTCGCGGATGCCGAAACCGGATGCCGAAACCGGG 135  
Db 766 CCGAGCGGCGCACCGGCACTCCGAGCGGTTCAGGAAACCGGAGTTGACCGCGCGCGCGC 707  
Qy 136 CCAGATCAGTGCAGCAGCGGATGCGCGCTCGATCTCGACGCTCGCGCGCAGCTCGA 195  
Db 706 CGATCCCGGGGAGGAGCGTGTGGCGGCGGACGCGGCGAATGACCGCCACCGACC 647  
Qy 196 TCTCGTCCCGGCTCCACACACAGGGGAAACTGGCTCGGTGGCAGCGGCGAGCCCGAC 255  
Db 646 GCCCGCGGCGAGCGGCTCGCCACCGCACCTGTGGCGAGCGCGCACACACCGAC 587  
Qy 256 GGGAGAGTGTCTCAGGTGGCGTTCAGGAGCTCTCGCGGTTCGGGAGCGGAGCGGAAACGCGCA 315  
Db 586 TCGACACCGCGCGCTCGCGGCTCCACAGAGCGCCGCTCGGCGCGCGAGCGCGT 527  
Qy 316 CGGCTCAGATCCCTGTCAGTGCGATCGGCTCAGTGCCTCGTCCCTTGGCCCTGGGAG 375  
Db 526 CGCGCGCGGCTCGATGACGCGGCCAGTGTTCATGGGCA--GAAAGACGGCGCGACGAG 469  
Qy 376 GATAGCGTTTACGACGAGCGCACCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 435  
Db 468 GACGAGCAGCGCGGAGCGCGGTGTCACGCGGAGCGGGTCCGTGCGGCGCTCGGCGCGCTC 409  
Qy 436 CGCTCGATGACGAGCGGCTCGGGGTTCGGGGCGGTTCGGCGAGATCCGTACCGCCCGGACC 495  
Db 408 CAGCGCTCGGGGGGTTCGGGGCGCTCGGGCGCGCTCGGGCGCGCTCGGGGGCGCGCGGGGC 349

Qy 496 GCCTCGGCGCAGG 508  
Db 348 GCCCGGGCGCGG 336  
RESULT 40  
US-10-156-761-7034/c  
; Sequence 7034, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7034  
; LENGTH: 2997  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(2997)  
US-10-156-761-7034  
Query Match 9.2%; Score 47.8; DB 15; Length 2997;  
Best Local Similarity 45.2%; Pred. No. 0.0074;  
Matches 214; Conservative 0; Mismatches 257; Indels 2; Gaps 1;  
Qy 1 GGTGAGCAGCAACGACCGGTGTTCGCCCCCGCGCCCATCAGCAATGCGCACTCGCGCG 60  
Db 1780 GCTCCAGCTCCGCTCCCTCGGCGAGGTGGCGCGGTTCGCGCGCGGTTCGCGCG 1721  
Qy 61 GGTGAGCAGCTCGCAGCAGGTTCGAGGAGTTCAGCAGCAGCTCATCCCGCCTCGCGCGCA 120  
Db 1720 TCGCGCGCGCAGCTTCCTCGCGTACGAGCGCGGTTCGCGCGCTTCGCGCGGTTCGCG 1661  
Qy 121 TCGCGGAAACACCGCGCGCAGGATCAGTGCAGCAGCGCGATGCGCGCTTCGATTCGACCG 180  
Db 1660 CGTGCGCGGTGTGGAGCGGTAGGCGGTAGGCGCGGTTCGCGCGCTTCGCGGTTCGACATGCCCG 1601  
Qy 181 TCGCGCGCAGCTCGATCTC--GTGCGCGGTTCGCCACACAGGCGGAAACTGGCTCGGTG 238  
Db 1600 TGATCTTCGCGCGCGCGCGCGGTTCGAGGAGCGCGATGTCGAGGAGCGCGCACACGCGCGAGCGCGCT 1541  
Qy 239 CAGCGCGACGCGCGCGGACAGCTTCGTCAGGCTGCGCTCAGCGACCTCTTCGCGGGTTC 298  
Db 1540 CGACGAGACCGCGCGCGAGTTCGCGCGGATGCCCTGAAGGCACTGTTCTTCAGGTCA 1481  
Qy 299 GGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTCTGATCGCATCGGCTCAGTCCCGGTCT 358  
Db 1480 GCCAGTGGACCGCGCGGTTCGCTCGCGCGCGCGCGAGGCGCACGCGCGCGCGCGCT 1421  
Qy 359 CCGCTTCGCTTCGGAGGATAGCGGTTTACGACGAGCGCGCACACGCGCGCGCGCGCGG 418  
Db 1420 CGTCCGTGCTTCGGAGAGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1361  
Qy 419 GGGCGGTTTTCAGCGGATTCGCTCGATGACAGCGGCTTCGCGGGGTTCGGGGCGGTTC 471  
Db 1360 GGTTCGCGCGCGCACGCGGAGCTCTCGCGCGCGGTTCGCGCGCTTCCTTCGCGCGAC 1308  
RESULT 41

US-10-156-761-3185/c  
; Sequence 3185, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 3185  
; LENGTH: 3789  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS (3789)  
; LOCATION: (1)..(3789)  
US-10-156-761-3185

Query Match 9.2%; Score 47.6; DB 15; Length 3789;  
Best Local Similarity 46.5%; Pred. No. 0.008;  
Matches 188; Conservative 0; Mismatches 214; Indels 2; Gaps 1;

QY	115	CGCGGATCCGGAACACCGGCGCAGATCATCGTGCACCGCGATCGCGCTCGATCT	174
Db	901	CGCGGAGTAGACCGCGCGCTTGGAGACCGCGCGCGCGACACCGAGGTGGGCT	842
QY	175	CGACGCTCGCGCGCAGCTCGATCTCGTCCCGCGCTCCACACCGAGGGAACCTGGCTCG	234
Db	841	CGCGCGGAGGACGAGCGGCGACCGGTTCGCGCGCGCGCGACCTCCCGCGCGAGCGGG	782
QY	235	GTGGCAGCGGACGCGCGCGCGGACGCTGCTCAGCGTGGCGTCCGCGACCTTCGCG	294
Db	781	CGCGGAGCAGCTCCGACCTGTGTGTCAGTTGCGGTAGCTCAGGACGCTGTCGCGGTGA	722
QY	295	GGTCCGCGGAGCGGAACCGCGCGCAGCTCGATCCCTGTGATCGCATCGGCTCAGTCCCGG	354
Db	721	CCAGCGCGGCTGTCCGGTGGTGGTGGCGCGCGCTCCGCGNACGATCTCTCGACGCGC	662
QY	355	TGCTCCCTTGGCTTGGAGGATAGCGGTTCCACGAGCGGCGACCGCGCGCGCGGG	414
Db	661	TGTCGCGCGGTAGGCGGACCGTCTGTCGCGGTTCAGTCCACCGAGTCCGCGGTTCGG	602
QY	415	GGGCGGCGGTTTCAGCCGAT--CCGCTCATACCGAGCGGCTCGCGGCTCGCGGCGGTTCG	472
Db	601	CGGCGGTTCAGCGGTTCAGTTCGCGACCGCGAGCTCGATCTGCCGCGAGGCGGTCA	542
QY	473	CGGAGTCCGTACCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGG	516
Db	541	GGCGCGCGGCGAGTGGAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	498

RESULT 42  
US-09-976-059-1/c  
; Sequence 1, Application US/09976059  
; Patent No. US2002016474A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Zazopoulos, Emmanuel  
; APPLICANT: Starfa, Alfredo  
; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
; FILE REFERENCE: 3019-PCT  
; CURRENT APPLICATION NUMBER: US/09/976,059

CURRENT FILING DATE: 2001-10-15  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 1  
LENGTH: 88421  
TYPE: DNA  
ORGANISM: Actinoplanes sp.  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (2077)..(3078)  
OTHER INFORMATION: ORF 1; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (3118)..(4032)  
OTHER INFORMATION: ORF 2; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (4038)..(5048)  
OTHER INFORMATION: ORF 3; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (6665)..(5814)  
OTHER INFORMATION: ORF 4; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (7703)..(6693)  
OTHER INFORMATION: ORF 5; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (9464)..(8130)  
OTHER INFORMATION: ORF 6; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (9691)..(10761)  
OTHER INFORMATION: ORF 7; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (12751)..(10829)  
OTHER INFORMATION: ORF 8; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (13617)..(12802)  
OTHER INFORMATION: ORF 9; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (15203)..(13614)  
OTHER INFORMATION: ORF 10; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (15591)..(15863)  
OTHER INFORMATION: ORF 11; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (15880)..(19035)  
OTHER INFORMATION: ORF 12; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (19032)..(39713)  
OTHER INFORMATION: ORF 13; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (39713)..(65800)  
OTHER INFORMATION: ORF 14; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (65826)..(65530)  
OTHER INFORMATION: ORF 15; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (66546)..(67370)  
OTHER INFORMATION: ORF 16; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (67384)..(70059)  
OTHER INFORMATION: ORF 17; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (70099)..(70662)  
OTHER INFORMATION: ORF 18; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (70659)..(71906)  
OTHER INFORMATION: ORF 19; positive strandedness  
NAME/KEY: misc feature  
LOCATION: (73439)..(71964)  
OTHER INFORMATION: ORF 20; negative strandedness  
NAME/KEY: misc feature  
LOCATION: (74216)..(73563)



```
RESULT 44
US-10-114-270-97/c
; Sequence 97, Application US/10114270
; Publication No. US20040030110A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Spathak, Kimberly A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Liu, Ziaohong
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Li, Li
; APPLICANT: Vernet, Corine
; APPLICANT: Zernusen, Bryan D.
; APPLICANT: Gorman, Linda
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Smithson, Glenna
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Gargolli, Esha A.
; APPLICANT: Taupier Jr., Raymond J.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Ji, Weizhen
; APPLICANT: Anderson, David W.
; APPLICANT: Lieté, Mario W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Stone, David J.
; APPLICANT: MacDougall, John R.
; APPLICANT: Rothenberg, Mark E.
; TITLE OF INVENTION: No. US20040030110A1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-322C
; CURRENT APPLICATION NUMBER: US/10/114,270
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/281,086
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,136
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/281,863
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/281,906
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/282,020
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/282,930
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,934
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/283,512
; PRIOR FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/283,710
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/284,234
; PRIOR FILING DATE: 2001-04-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SEQ ID NO 97
; LENGTH: 5571
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS:
; LOCATION: (67)..(5494)
US-10-114-270-97
Query Match 9.1%; Score 47.2; DB 13; Length 5571;
Best Local Similarity 47.1%; Pred. No. 0.0094;
Matches 178; Conservative 0; Mismatches 198; Indels 2; Gaps 1;
```

## RESULT 45

```
US-10-282-122A-30147/c
; Sequence 30147, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```



```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30147
; LENGTH: 1611
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30147

Query Match          9.0%; Score 47; DB 13; Length 1611;
Best Local Similarity 52.2%; Pred. No. 0.013;
Matches 132; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

Qy 65 GACCAGTTCAGCAGGAGTGGAGTTAGACAGCCAGCTCATCCGCGCTCCGCGCGGATGCC 124
Db 660 GACCAGGTTGAGTTCGTTGGTGGTATCAGCAGCAGGAGATGCCAGGCGTGCCTGCAGTC 601
Qy 125 GAAACACCGCGCCAGGATCAGTCGACGACGAGCGATCGCGCTTCGATCTCGACGTCGG 184
Db 600 CTTGAGCAGATCGAGGATCTTCACTGACCGTGA-----CGTCGAGCGGTGTCGG 547
Qy 185 CCGCAGCTCGATCTGTCGCCCGCTCCACACAGGAGGAACTGGCTCGTGGCAGCGG 244
Db 546 TTCGTGCGCGATCAGCAGTTTCGGCTCGCAGCGCCAGGCGCATGGCGATCATCAGCGCTG 487
Qy 245 CAGCCCCAGCGGACAGCTCGTCCAGCTGGCTCAGGACTCTCGCGGCTCGGGAGC 304
Db 486 CCGCTGCGCGCGGAGAGTTCGTGTGATAGGCTTTCAGGCGCTTCGCGGCTCGGGGAT 427
Qy 305 GGAACAGCGCAGC 317
Db 426 GCCGACCAGCTCG 414

RESULT 46
US-09-779-429-2/c
; Sequence 2, Application US/09779429
; Patent No. US20010007156A1
; GENERAL INFORMATION:
; APPLICANT: Duivick, John
; APPLICANT: Gilliam, Jacob
; TITLE OF INVENTION: A Hydroperoxide Lyase Gene from Maize and Methods of
; FILE REFERENCE: Maize hydroperoxide lyase (HPL)
; CURRENT APPLICATION NUMBER: US/09779,429
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 09/417,704
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1835
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(115)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1025)..(1027)
; OTHER INFORMATION: I-helix
; NAME/KEY: misc binding
; LOCATION: (1457)..(1459)
; OTHER INFORMATION: Heme-binding site
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1625)..(1835)
; US-09-779-429-2

Query Match          9.0%; Score 47; DB 9; Length 1835;
Best Local Similarity 50.2%; Pred. No. 0.012;
Matches 116; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 20 GTGGTCCGCGCGCCCATCAGAACTGCCATCTCCGCGGGGTGACAGGTGCACCG 79
Db 1528 GTCGTGTCGAGCGCGGAAACAGTCGCGCCAGCAGCATGCGCGCTGTCCACCCACCTC 1469
Qy 80 GTGGGAGTTGAGCAGCAGCTCATCGCCCTGCGCGCGCATGCGGAAACACCGGGCCAG 139
Db 1468 CTTGCGCGCGCACTGCTTTGTTCCCGGGCCCGGGCTCTCCGGCCCGCTTGGACCA 1409
Qy 140 GATCAGTTCAGCAGCGCGATGCGGCTTCGATCTCGACGCTCGGCGCGAGTCTCGATCTC 199
Db 1408 GAAGAGTGTCTGACGAGCGCTTCGCGCGCTTCGTCGCGGAGGAGCGTTCCGGCAGCACTC 1349
Qy 200 GTCGCCCGGCTCCACACAGGAGAACTGGCTCGGTGGCAGCGCAGCCC 250
Db 1348 CTCGGCGCGCTCGAACAACCTCGGGTCCCGATCGCCAGCGGCTGGTACCC 1298

RESULT 47
US-10-193-748-2/c
; Sequence 2, Application US/10193748
; Publication No. US20030167510A1
; GENERAL INFORMATION:
; APPLICANT: Duivick, John
; APPLICANT: Gilliam, Jacob
; TITLE OF INVENTION: A Hydroperoxide Lyase Gene from Maize and Methods of
; FILE REFERENCE: Maize hydroperoxide lyase (HPL)
; CURRENT APPLICATION NUMBER: US/10193,748
; PRIOR FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: US/09/417,704
; PRIOR FILING DATE: 1999-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1835
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(115)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1025)..(1027)
; OTHER INFORMATION: I-helix
; NAME/KEY: misc binding
; LOCATION: (1457)..(1459)
; OTHER INFORMATION: Heme-binding site
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1625)..(1835)
; US-10-193-748-2

Query Match          9.0%; Score 47; DB 15; Length 1835;
Best Local Similarity 50.2%; Pred. No. 0.012;
Matches 116; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

Qy 20 GTGGTCCGCGCGCCCATCAGAACTGCCATCTCCGCGGGGTGACAGGTGCACCG 79
Db 1528 GTCGTGTCGAGCGCGGAAACAGTCGCGCCAGCAGCATGCGCGCTGTCCACCCACCTC 1469
Qy 80 GTGGGAGTTGAGCAGCAGCTCATCGCCCTGCGCGCGCATGCGGAAACACCGGGCCAG 139
Db 1468 CTTGCGCGCGCACTGCTTTGTTCCCGGGCCCGGGCTCTCCGGCCCGCTTGGACCA 1409
Qy 140 GATCAGTTCAGCAGCGCGATGCGGCTTCGATCTCGACGCTCGGCGCGAGTCTCGATCTC 199
Db 1408 GAAGAGTGTCTGACGAGCGCTTCGCGCGCTTCGTCGCGGAGGAGCGTTCCGGCAGCACTC 1349
Qy 200 GTCGCCCGGCTCCACACAGGAGAACTGGCTCGGTGGCAGCGCAGCCC 250
Db 1348 CTCGGCGCGCTCGAACAACCTCGGGTCCCGATCGCCAGCGGCTGGTACCC 1298

RESULT 48
US-10-156-761-6368/c
; Sequence 6368, Application US/10156761
```



```
; PRIOR APPLICATION NUMBER: US 60/224,729
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 38
; LENGTH: 7789
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7477141CB1
US-10-333-314-38
```

```
Query Match          9.0%; Score 46.8; DB 13; Length 7789;
Best Local Similarity 50.0%; Pred. No. 0.011;
Matches 117; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

Qy 29 CCGCCCGCCCTCAGCACTCGCGGGGTGACCAAGTCCGACGAGGTGGAGTT 89
Db 4589 CCGTCCGAACAGCGCGGTGCTGCGGCCGACGCGCCCGAGTCTCTGCTGCCACT 4530
Qy 89 GAGCAGCCAGCTCATCGCGGCTCGCGGCGATGCCGAAACACCGGSCCAGGATCAGTG 148
Db 4529 GCGTGCATATCGGTGACAGCCGCTCAGGGTGAAGCTCAGCGCCTGCGCATGCCAG 4470
Qy 149 CAGCAGCGCGATGCGCGCTCGATCTCAGCGTCCGCGCGAGTCTGATCTCGTCCCGG 208
Db 4469 CACCGGGGAGCCCGCGCGAGCTCCGCGCCTCCGAGTCTCTCCGTCCCGCGCGGGC 4410
Qy 209 CTCCACACACAGGGGAACTGGCTCGGTGGCAGCGCGAGCCCGAGCGGGACAG 262
Db 4409 CTCCAGCGCGGTGGCGCTCCGCGGAGGGGTCCCGCGAGCGCGCTGGGACAG 4356
```

Search completed: June 27, 2004, 21:03:00  
Job time : 374.196 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:12 ; Search time 2109.57 Seconds  
(without alignments)  
7360.905 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_109000\_109519  
Perfect score: 520  
Sequence: 1 ggtgcagcacgaacg...ggccagcgccggggaacc 520

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : EST:

```

1: em_estba:*
2: em_estum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_estfun:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_sss1:*
29: gb_sss2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	76	14.6	925	29	CNS0091P
C 2	75	14.4	982	13	BX415111
C 3	68.8	13.2	925	29	CNS0091P
C 4	67.6	13.0	924	13	BX442207

#### RESULT 1

CNS0091P/c  
LOCUS

DEFINITION

BACR19D16 of RPEC1-98 library from Drosophila melanogaster (fruit fly); genomic survey sequence.

ACCESSION

AL053013

VERSION

AL053013.1

KEYWORDS

GSS.

SOURCE

Drosophila melanogaster (fruit fly)

ORGANISM

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 925)

Genoscope.

TITLE

Direct Submission

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

Web : www.genoscope.cns.fr

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

#### ALIGNMENTS

CNS0091P 925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPEC1-98 library from Drosophila melanogaster (fruit fly); genomic survey sequence.

AL053013 GI:4934461  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 925)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

5	65	12.5	982	13	BX415111
6	64.2	12.3	932	29	CNS0072Q
7	64	12.3	1288	13	BQ578719
8	63.4	12.2	1153	29	AG065289
9	62.2	12.0	1124	29	AG041123
C 10	61.4	11.8	91	29	CG243505
C 11	61.4	11.8	932	29	CNS0072Q
12	60.2	11.6	844	29	AG083352
C 13	59.6	11.5	1409	12	BG850730
C 14	59.2	11.4	924	13	BX442207
15	58.8	11.3	767	29	AG076679
16	58.8	11.3	835	14	CB904973
17	58.8	11.3	935	14	CF876206
C 18	58.6	11.3	925	29	CNS0065X
C 19	58.6	11.3	1160	13	BUS37937
C 20	58.4	11.2	935	29	CNS0065X
C 21	58	11.2	1015	28	BZ569259
22	58	11.2	1221	29	AG058113
23	57.8	11.1	693	29	AG166257
24	57.8	11.1	878	12	BG785566
C 25	57.8	11.1	1189	29	AG032118
26	57.4	11.0	1387	10	AW731151
27	57	11.0	1201	13	BX405071
28	57	11.0	1785	28	CC219595
29	56.8	10.9	1085	29	AG050224
C 30	56.8	10.9	1610	28	BZ569386
C 31	56.6	10.9	483	9	AJ442101
C 32	56.6	10.9	593	29	CG465713
C 33	56.4	10.8	976	29	AG041117
34	56.2	10.8	978	29	AG042900
C 35	56	10.8	1203	29	CNS015Y4
C 36	56	10.8	1449	12	BM912953
37	55.8	10.7	851	29	AG171188
38	55.8	10.7	1273	12	BM562099
39	55.6	10.7	1094	14	CK161715
C 40	55.6	10.7	1203	29	CNS015Y4
41	55.6	10.7	1348	29	CG752544
42	55.4	10.7	818	14	CK200413
C 43	55.4	10.7	1086	29	AG077581
C 44	55.4	10.7	1232	29	AG072425
C 45	55.2	10.6	608	28	BZ639482
C 46	55.2	10.6	633	29	CG623358
C 47	55.2	10.6	912	29	CG328665
C 48	55.2	10.6	1273	12	BM562099
C 49	55	10.6	734	13	BU400981
C 50	55	10.6	1285	29	AG058001





```

ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished (2001)
COMMENT        Contact: Genoscope
                Genoscope - Centre National de Sequencage
                BP 191 91006 EVRY cedex - France
                Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
                Library was constructed by Life Technologies, a division of
                Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
                http://fulllength.invitrogen.com/Invitrogen Corporation 1600
                Faraday Avenue Genoscope sequence ID : CSOCAP004BE10QPI.
                Location/Qualifiers
                1..982
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSOCAP004Y120"
                /tissue_type="THYMUS"
                /clone_lib="Homo sapiens THYMUS"
                /note="Vector; pCMVSPORT 6; 1st strand cDNA was primed
                with a NotI-oligo(dT) primer. Five prime end enriched,
                double-strand cDNA was digested with Not I and cloned into
                the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                Library was not normalized."

FEATURES             source
    source
    Query Match      12.5%; Score 65; DB 13; Length 982;
    Best Local Similarity 29.5%; Pred. No. 0.85;
    Matches 129; Conservative 102; Mismatches 206; Indels 0; Gaps 0;

QY 79 GGTGGAGTTAGCAGCAGCTATCCCGCTCGCGCGGATGCGGAACACACGGGCCA 138
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 GGGGGGSSSGVGGVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 181

QY 139 GGATCAGTCAGCAGCGGATGCGCGCTCGATCTCGACGGTCGCGCGCAGCTCGATCT 198
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 CCSSGKCCCGKCCCCCGGGGGGCCCCSSSSSSSSSSSSSSSSSSSSSSSSSSSS 241

QY 199 GGTGCGCCGGTCCGACACAGAGGGAACCTGGTCTGGTGGCAGCGGACCCAGCGGG 258
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 CCCCCCCCCCGCCCCCGGGGCCCCCGCCSSSSSSSSSSSSSSSSSSSSSSSSSSSS 301

QY 259 ACAGTCTGTCAGGTGGCGTCACGGACCTCTCGCGGTTCGGGAGCGGAACGCGACGG 318
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 GSGGGGCGSSSCGSGSGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 361

QY 319 CTCAGATCCCTGTAGTCGATCGATCGGCTCAGTCGCGGTCTGTCCTTGGCTGGAGGAT 378
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 GGGSSSSSCGSSSCSSSGGGSSSSSGGGSSSGGGSSSSSSSSSSSSSSSSSSSSSS 421

QY 379 AGCGTTTCACGACGCGCACACGCGCGGCGGGGCGGGGCGGTTCAGCCGATCCG 438
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 422 GSSGSGSSGSGSGSGSGSGSGCGCGCGSGGGGSGGGGSGGGGSGGGGSGGGGSGGG 481

QY 439 TCAGTACACGCGGTTCGGGGTTCGGGCGGTTCGGGCGGATCCGTCACGCGCGACCGCC 498
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 GSGCGGGGGGCGSSCGSSGGGGGGGGGGGGSSCGSSCGSCCGGCCCCCGGCGCGSC 541

QY 499 TGGGCGACCGCGCGG 515
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 CCCCCCGGCCCCCAG 558

RESULT 6
CNS00720
LOCUS      CNS00720      932 bp      DNA      linear      GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #
            BAC14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.

```

```

AL066742
AL066742.1 GI:4945205
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Ephydroidea; Drosophilidae; Drosophila.
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Oseegawa and
Aaron Mammoss in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..932
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BAC14B09"
/clone_lib="RPCI-98"
/note="end : T7"

FEATURES             source
    source
    Query Match      12.3%; Score 64.2; DB 29; Length 932;
    Best Local Similarity 30.9%; Pred. No. 1.1;
    Matches 109; Conservative 86; Mismatches 158; Indels 0; Gaps 0;

QY 161 GCGCGCTCGATCTCGACGGTCGCGCGGAGTCGATCTCGCGCGGCTCCACACGAG 220
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 555 GSKSGSGSGSGSBGCGGGGKTGSGGGGGSBTKYTBKGGGCGSGSSCCSGSGSRG 614

QY 221 GGGAACTGGCTCGGTGCGCGGCGGAGCCCGCGCGCGCGCGCGCGCGCGCGCG 280
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 615 GGGGCGSGSGSGCGCGCGGSGSGSGCGSGCGCGCGCGCGCGCGCGCGCGCGSG 674

QY 281 ACGGACCTTTCGCGGGTCGGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTGATCGAT 340
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 675 GCGGCGSSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 734

QY 341 CGGCTCAGTTCGCTCGTCCCTTGGCTGGGAGTAGCGGTTACAGAGCGGCGAC 400
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 735 CCCCCCCCCSSCCSSSSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 794

QY 401 CACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 460
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 795 CGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 854

QY 461 TCGGGGCGGTTCGCGAGATCCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 855 SSGCGSGSGSGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGSC 907

RESULT 7
CNS00720
LOCUS      CNS00720      1288 bp      mRNA      linear      EST 15-JUL-2002
DEFINITION AGENCOURT_819257 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:621563
            5', mRNA sequence.

```

```

ACCESSION      BQ678719
VERSION        BQ678719.1  GI:21791398
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1288)
AUTHORS        NIH-MGC http://mgi.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-remail.nih.gov
               Tissue Procurement: DCTD/BTP
               cDNA Library Preparation: Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMML)
               DNA Sequencing by: Agencourt Bioscience Corporation
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLML at:
               http://image.llnl.gov
               Plate: LCM2422 row: h column: 12
               High quality sequence stop: 171.

FEATURES       source
               1..1288
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:6261563"
               /tissue_type="melanotic melanoma, cell line"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH MGC 112"
               /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
               EcoRI; cDNA made by oligo-dT priming. Directionally cloned
               into EcoRI/XhoI sites using the following 5' adaptor:
               GGCACGAG(G). Library constructed by Ling Hong in the
               laboratory of Gerald M. Rubin (University of California,
               Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
               Superscript II RT (Life Technologies). Note: this is a
               NIH_MGC Library."

ORIGIN
Query Match      12.3%; Score 64; DB 13; Length 1288;
Best Local Similarity 42.8%; Pred. No. 1.3; Mismatches 0; Gaps 0;
Matches 205; Conservative 0; Indels 274; Indels 0; Gaps 0;

QY 22 GGTCCGCCCGCGCGCCATCAGAACTCCACTCGCGCGGGTGACCAAGTFCGACCAAGGT 81
DB 498 GGGCGCGCGCGCGCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547

QY 82 GGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGCGCATCGCAACACCGCGCCAGGA 141
DB 548 GCGGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607

QY 142 TCACGTGCAGCAGCGCGATGCGCGCGCTCGATCTCGACGCTCGCGCGCGCGCGCGTCTCGT 201
DB 608 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 667

QY 202 CGCCCGGCTCCACACAGGGGAACTGGCTGCTGGCAGCGGACGCCAGCCAGCGGACACA 261
DB 668 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 727

QY 262 GCTCGTCAGAGTGGCGTCACGAGCTCTCGCGGCTCGGAGCGGAAACGGCGACGGCTC 321
DB 728 NCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 787

QY 322 AGATCCCTGTGAGTCAGTCAGGCTCAGTGCAGGCTCGTCCCGCTTGGCTGGAGGATAGC 381
DB 788 CNGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 847

QY 382 GTTTCACGACGAGCGGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 441
DB 848 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 907

```

```

QY 442 ATGACACAGCGGCTCGGGGTGCGGGGTCGCGGAGATCCGTACCGCGCGGACCGGCTC 500
DB 908 CMC CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966

RESULT 8
AG065289      1153 bp      DNA      linear      GSS 03-NOV-2001
LOCUS         Pan troglodytes DNA, clone: PTB-054K01.F, genomic survey sequence.
DEFINITION    AG065289
ACCESSION     AG065289
VERSION       AG065289.1  GI:16617091
KEYWORDS      GSS.
SOURCE        Pan troglodytes (chimpanzee)
ORGANISM      Pan troglodytes
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE      1
AUTHORS        Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
               Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE          BAC end sequences of Library PTB
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 1153)
AUTHORS        Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
               Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE          Direct Submission
JOURNAL        Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
               (E-mail: chimpanse@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/;
               Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT        Clones are derived from the chimpanzee BAC library PTB This BAC end
               was generated during the R&D process and may have higher chance of
               clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..1153
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-054K01.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match      12.2%; Score 63.4; DB 29; Length 1153;
Best Local Similarity 46.3%; Pred. No. 1.6; Mismatches 277; Indels 1; Gaps 1;
Matches 240; Conservative 0;

QY 4 GCAGCACGAAACGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 63
DB 211 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270

QY 64 TGACCAAGGTTCAGCAGCGTGGAGTTGAGCAGCGCAGCTCATCGCGCGCTCGCGCGGATGC 123
DB 271 CGGCGACGACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAG 330

QY 124 CGAAACACCGCGCGCGATCATGTGACGACGCGCGCGCGCGCGCTCGATCTCGACGGTCG 183
DB 331 CGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390

QY 184 GCGCGAGCTCGATCTCGTCCCGCGGTCTCCACACAGGGGAACTGGCTCGGTGGCAGCG 243
DB 391 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 450

QY 244 GCAGCCCCAGCGCGGACGCTGTCAGGCTGCGCTCAGGACCTCTCGCGGCTCGGGAG 303
DB 451 CGGCTAGCGCGCGAGGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510

```





Matches 232; Conservative 0; Mismatches 266; Indels 1; Gaps 1;

QY 22 GGTGCGCCCGCGCCATCAGAACTGCACTCCGCGGGGTGACCAAGGTGCGACAGGT 81

Db 503 GTCTCCCTCTGTGCTTGTGACACGAGGCTCTGTCAGGCTCTTGGACCCGCGGC 444

QY 82 GGGAGTTGAGCAGCAGCTCATGCGCGCTGCGCGGATGCGGAAACACCGGCGCAGGA 141

Db 443 CCCCCTCCGCGCGCGCGCGCGAGGCGCGCGCGCGCGATGCGGAACTGCGGCTGCGGA 384

QY 142 TCACGTGACGACGCGCGATGCGCGCTCGATCTCAGCGTGGCGCGAGCTCGATCTCGT 201

Db 383 CCAGAGCAGATTGTCGTTGAACCGAGTCCGCCACCGTGGCGCGCGCTCTCTCAGG 324

QY 202 CCGCCGCTCCACACCGAGGGAATCGCTCGTGGCAGCGCAGCCCGCGCGGACAC 261

Db 323 TGAGACCGACGTCGCGCGTCAAGAGCGCGCGCTCGCTCGCGCTCTCTCAGAGCTCC 264

QY 262 GCTCTCCAGGTGCGCTCAGGACCTCTCGCGGTGCGGAGCGGAAACCGGCGCGCTC 321

Db 263 TCTCCACTGTCGCGCGCCAGTCTCTCGCGCTCGCGCGGAGTGGCGCCACACCG 204

QY 322 AGATCCCTGTAGTCCGATCGCTCAGTCCGCTCGTCTCCCTCGCTGGCGAGGATAGC 381

Db 203 GGACAGCTCCACG 144

QY 382 GCTTCAAGCAGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 441

Db 143 CTTCCAGAGGACCACTCCAGCGCTCCCGCGCTCGCGCTCGATCTT-GTCCCGCTG 85

QY 442 ATGACAGCGCTGCGCGGTGCGCGGTGCGCGAGATCCGTACCGCGCGCGCGCGCTCG 501

Db 84 AGATCGGCGAGTGGCGCGCGCTCGGAAACCGCAGCTTCCAGCGCGCGCGCTCGCGCC 25

QY 502 GCAGGCGCGCGGATCC 520

Db 24 GCGCTCTCGCGCGCGCGCC 6

RESULT 11

LOCUS CNS00720/c

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BAC14809 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066742

VERSION AL066742.1 GI:4945205

KEYWORDS GSS

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 932)

AUTHORS Genoscope

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES

Source

Location/Qualifiers

1..932

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BAC14809"

/clone\_lib="RPCI-98"

/notes="end : T7"

ORIGIN

Query Match 11.8%; Score 61.4; DB 29; Length 932;

Best Local Similarity 32.0%; Pred. No. 3.1;

Matches 108; Conservative 76; Mismatches 153; Indels 0; Gaps 0;

QY 169 CGATCTCAGCGTCGCGCGAGCTCGATCTCGTCGCGCGCTCCACACCGAGGAACT 228

Db 929 CG 870

QY 229 GGTCTGTCGACGCGCGAGCCCGCGCGGACAGCTGTCACAGCTGGCGTTCACGACCT 288

Db 869 GSCG 810

QY 289 CTGCGCGGTGCGGAGCGGAAACGCGCAGCTCAGATCCCTGTCAGTTCGATCGGCTCAG 348

Db 809 SSSCG 750

QY 349 TGCCTGTCGTCCTTGGCTTGGCGGAGATACGCGTTTACGAGCGGCGCACCGCGCG 408

Db 749 CGSAGCG 690

QY 409 GCGGCGCGCGCGCGCGCTTACGCGATCGCTCGATCAGCAGCGCTGCGGCTCGGCGCG 468

Db 689 GCSGCG 630

QY 469 GTCGCGGAGATCGTACCGCGCGCGCGCGCGCGCTCGGCGCA 505

Db 629 CGCCCG 593

RESULT 12

LOCUS AG083352

DEFINITION Pan troglodytes DNA, clone: PTB-080M08.R, genomic survey sequence.

ACCESSION AG083352

VERSION AG083352.1 GI:16635154

KEYWORDS GSS

SOURCE Pan troglodytes (chimpanzee)

ORGANISM Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE 1

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

TITLE BAC end sequences of Library PTB

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 844)

TITLE Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.

JOURNAL Direct Submission

COMMENT Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@sc.riken.go.jp, URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: M13Rev

LIBRARY Vector : pKS145

R.Site 1 : SacI

```

FEATURES
  Source
    Location/Qualifiers
      1..844
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-080M08.R"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
  Query Match      11.6%; Score 60.2; DB 29; Length 844;
  Best Local Similarity 46.1%; Pred. No. 4.7;
  Matches 211; Conservative 0; Mismatches 246; Indels 1; Gaps 1;

QY 53 CTCGGGGGGTGACAGGTGACAGGTGGAGTTAGACAGCAGTCTATCCCGCTG 112
Db 282 CNGCCGCGCNCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 341
QY 113 CGCGCGATGCCGAAACACCGGGCAGGATCAGTGCAGACACGCGGATGCGCCTCGAT 172
Db 342 CGGGCGCGCGCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401
QY 173 CTCGACGGTC-GGCGCAGCTCGATCTCGTCGCCGCTCCACACACGAGGAACTGGC 231
Db 402 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461
QY 232 TCGGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 291
Db 462 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
QY 292 CGCGGTGGGAGCGGAAACGCGGACGCTCAGATCCCTGTCATCGATCGATCGATG 351
Db 522 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
QY 352 CGGTGCTCCCTTGGCTGGAGATAGCGGTTACAGACGAGCGGACACCGCGCGG 411
Db 582 GGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 641
QY 412 GGGCGGGGGGGGTTACGCGATCCGCTCGATGACGAGCGGTGCGGGGTTGCGGTC 471
Db 642 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 701
QY 472 GCGGAGATCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 509
Db 702 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 739

RESULT 13
BG850730      1409 bp      mRNA      linear      EST 29-MAY-2001
LOCUS      1024029B11.v1 C. reinhardtii CC-1690, normalized, Lambda Zap II
DEFINITION      Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION      BG850730
VERSION      BG850730.1 GI:14231914
KEYWORDS      EST.
SOURCE      Chlamydomonas reinhardtii
ORGANISM      Chlamydomonas reinhardtii
              Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
              Chlamydomonadales; Chlamydomonas.
              1 (bases 1 to 1409)
REFERENCE
  AUTHORS      Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
              McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE      Analyses of the Chlamydomonas reinhardtii Genome: A Model,
              Unicellular System for Analyzing Gene Function and Regulation in
              Vascular Plants; project phase 2
  JOURNAL      Unpublished (2000)
  COMMENT      Contact: Charles Hauser
              DCM Box 91000
              Duke University
              Durham, NC 27708-1000
              Tel: 919 613 8159

FEATURES
  Source
    Location/Qualifiers
      1..1409
        /organism="Chlamydomonas reinhardtii"
        /mol_type="mRNA"
        /strain="CC-1690 wild type mt+ 21gr"
        /db_xref="taxon:3055"
        /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
        II"
        /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:
        XhoI; This library, constructed by John Davies and Jeffrey
        McDermott, combines cDNAs from CC-1690 cells grown in the
        mid-log phase in TAP (acetate-containing) medium in the
        light, TAP medium in the dark, HS (minimal) medium in
        ambient levels of CO2 and HS medium bubbled with 5% CO2.
        PolyA mRNA was purified from each sample, pooled and cDNA
        synthesized. The cDNA was directionally cloned into lambda
        ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
        pBluescript II SK- plasmids were excised from the lambda
        ZAP clones by superinfection with ExAssist (Stratagene)
        phage. The library was normalized using method 4 described
        in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN
  Query Match      11.5%; Score 59.6; DB 12; Length 1409;
  Best Local Similarity 44.3%; Pred. No. 6.5;
  Matches 218; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

QY 25 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 84
Db 376 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435
QY 85 AGTTGAGCAGCAGGTCTCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 144
Db 436 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495
QY 145 GTGACAGCAGCGCATGCGCGCTCGATCTCGAGCTGCGCGCGCGCGCGCG 204
Db 496 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 555
QY 205 CGGCTCCCAACACAGGGGAACTGGCTCGGTGCGCGCGCGCGCGCGCGCG 264
Db 556 AGAGACAGCAGCGGTTGTTGTTGCGCGCGCGCGCGCGCGCGCGCGCGCG 615
QY 265 GTCCAGCTGGGTACGAGACCTCTCGCGCGCGCGCGCGCGCGCGCGCGCG 324
Db 616 CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 675
QY 325 TCCCTGTGATCGCATCGGTCTAGTGCCTGCTCCCTTTGGCTCGGAGGATAGCGGT 384
Db 676 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 735
QY 385 TCAGACAGCGCGCACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444
Db 736 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 795
QY 445 ACCAGCGCTCGGGGTGCGGGCGGTCTCGCGAGATCCGTACCGCGCGCGCG 504
Db 796 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 855
QY 505 AGCGCGCGCGCGG 516
Db 856 GCGCGCGCGCGG 867

RESULT 14
BX442207/c
LOCUS      BX442207 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION      CS0DF027YD08 5-PRIME, mRNA sequence.
ACCESSION      BX442207
VERSION      BX442207.1 GI:30786042
```

```

EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 924)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF027DB04Qp1.

FEATURES
            Location/Qualifiers
                1..924
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="CS0DF027YD08"
                    /tissue_type="FETAL BRAIN"
                    /dev_stage="fetal"
                    /clone_lib="Homo sapiens FETAL BRAIN"
                    /note="Torgan: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                    was primed with a NotI-oligo (dT) primer. Five prime end
                    enriched, double-strand cDNA was digested with Not I and
                    cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                    vector. Library was not normalized."

ORIGIN
Query Match      11.4%; Score 59.2; DB 13; Length 924;
Best Local Similarity 36.6%; Pred. No. 6.9;
Matches 138; Conservative 53; Mismatches 186; Indels 0; Gaps 0

QY 134  GGCACGAGTACGTGCAGCAGCGGCGATCGCGGCTCTCGATCTCACGCTCGGCGCAGCTC 193
DB 913  GCGGSGCGSSGGCGGCGCGCGGSGGSGCGSCCGCCCGCGGGGCGNCCCGCGCGC 854
QY 194  GATCTCGTCCCGCGGCTCCACACAGGGGAAAATGGCTCGTGGCAGGGGAGCCCCAG 253
DB 853  CGCCSCGCGCGNCGGSGSGGCGGCGCCGCCCGCGGSGGGGCGGCGSCCGCCGCC 794
QY 254  CGGGACAGCTCGTACAGCTGGGTACAGACCTCTCGGGGTTCGGAGCGGAACGG 313
DB 793  GGGGSCSCGCGSGGCGCGCCCGCCCGCGGCGCCSCCGGCGGCGCGCGCGCGCG 734
QY 314  CACGCTCAGATCCTCTGTTCAGTTCGATCGGTTCAGTCCCGTCTGTCCTCCCTTGGCTGG 373
DB 733  SSSGCGSCCGCGCTKGCSCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 674
QY 374  AGGATAGCGTTACAGACAGGGGACACACGCGCGGCGGGGCGGGGCGGGTTCAGCGA 433
DB 673  SCGCGGGGGCGCCCGCGSGGGGGGGGCGCCGCCSCGCGCGGCGCGCGCGCGCGCG 614
QY 434  TCCGCTTCAGTACACAGCGGCTTCGGGGGTTCGGGGCGGTTCGCGCAGATCCGTTACGCCGA 493
DB 613  SSGCCCGGSGSGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 554
QY 494  CGGCTTCGCCACAGCGCC 510
DB 553  CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537

RESULT 15
AG076679
LOCUS
DEFINITION
AG076679
ACCESSION
AG076679.1 GI:16628481
VERSION
AG076679.1
KEYWORDS
GSS.

```

**SOURCE**  
**ORGANISM**  
Pan troglodytes (chimpanzee)  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
1  
**REFERENCE**  
**AUTHORS**  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
BAC end sequences of Library PTB  
Unpublished  
2 (bases 1 to 767)  
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
Totoki, Y., Watanabe, H. and Sakaki, Y.  
Direct Submission  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp-gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
Clones are derived from the chimpanzee BAC library PTB. This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
**PRIMERS**  
Sequencing: -21M13  
**LIBRARY**  
Vector : pKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
Location/Qualifiers  
1. 767  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-070P08.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"  
**FEATURES**  
**source**  
Query Match 11.3%; Score 58.8; DB 29; Length 767;  
Best Local Similarity 47.1%; Pred. No. 7; 2;  
Matches 236; Conservative 0; Mismatches 256; Indels 9; Gaps 2;  
QY 25 CGCCCCGGCGGCCCATCATCAAGATGCGATCTCGCGCGGGGTGACACAGTGCACAGGTGGG 84  
DB 87 CCCCCCGCGCCCCCCCCCGCGCGCGCCCCCNCCNCCNCCNCCNCCNCCNCCGCGCCCCCCCCCGCC 146  
QY 85 AGTTGAGCAGCCAGCTCATCGCGCTCGCGCGGCATGCCGAACACCGGGCCAGCATCA 144  
DB 147 CGGGGCGCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCGCGCGCC 206  
QY 145 CGTGCAGCAGCGCGATGCGCGCTTCGATCTCAGCGTGGCGCGCAGCTCGATCTCGTCGC 204  
DB 207 GCCCCCCCCCCCCCCCCCCAACCCCGTCACTGCCCGCGCGCGCCCCCGCCCCCCCCCGC 266  
QY 205 CC-----GGCTCCACACACAGGGGAAATGSGCTCGTGGCAGCGCAGCCCGAGCGGG 258  
DB 267 CC CGCGGGGCGCCCCCGCGCGCGCCCCCGGGGCCCCCGCGCGCGGGGGGGCCCCCGCGCGCG 326  
QY 259 ACAGCTCTGCTCAGGCTTGGGGTCAACGACCTCTCG---CGGG"CGGGAGCGGAAACCGCA 315  
DB 327 GGGCGCGCCCCCGCGCGCGCGCGGGGNGCCCCCGCGCGCGGGGGCGCGCGGCGCCCG 386  
QY 316 CGGCTCAGATCCCTGTCACTGCGATCGGCTCAAGTCCGCTGCTGTCCTTGGCTGGAG 375  
DB 387 CGCGCGCGCGCGCGCGCCCCCGCCCCCGCGCGCCCCCGCGCGCGCGCGCTGCTGCCCGCGCC 446  
QY 376 GATAGCGGTTCACGACGAGCGGCACCAACCGCGCGGGGGGGCGGGGGCGGTTCAGCCGATC 435  
DB 447 GCGCGGGGGCGCGCCCGGGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 506  
QY 436 CGCTCGATGACGACGCGCTGCGGGTTCGGGGCGGTTCGGCGAGATCGATCCGCCCGCGAC 495  
DB 507 CACCGGGGGCG 566

RESULT 15  
AG076679 LOCUS  
AG076679 767 bp DNA linear GSS 03-NOV-2001  
DEFINITION Pan troglodytes DNA, clone: PTB-070P08.F, genomic survey sequence.  
ACCESSION AG076679  
VERSION AG076679.1 GI:16628481  
KEYWORDS GSS.

[illegible]

```

QY 496 GCCTCGGCCAGCGCCGCGGG 516
Db 567 CGGCGCGCGCGCGCGGG 587

RESULT 16
CB904973 835 bp mRNA linear EST 02-JUL-2003
LOCUS trico40xn11 T.reesei mycelial culture, Version 3 april Hypocrea
DEFINITION jecorina cDNA clone trico40xn11, mRNA sequence.
ACCESSION CB904973
VERSION CB904973.1 GI:301119631
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
REFERENCE Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
AUTHORS Foreman,P.K., Brown,D.E., Dankmeyer,L., Dean,R., Diener,S.,
Dunn-Coleman,N.S., Goedegebuur,F., Houfek,T.D., England,G.J.,
Kelley,A.S., Meerman,H.J., Mitchell,T., Mitchinson,C.,
Olivares,H.A., Teunissen,P.J., Yao,J. and Ward,M.
TITLE Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
JOURNAL J. Biol. Chem. 278 (34), 31988-31997 (2003)
MEDLINE 22803314
PUBMED 12789320
COMMENT Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
Location/Qualifiers
1. 835
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico40xn11"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 3 april"
/note="Vector: pREP3y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match 11.3%; Score 58.8; DB 14; Length 835;
Best Local Similarity 46.7%; Pred. No. 7.9;
Matches 243; Conservative 0; Mismatches 273; Indels 4; Gaps 2;

QY 2 GTGCAGCAGCAACGCGGTGGTCCGCCGCCGCCATCACGAATGCCACTCCGCGG 61
Db 299 GGGCGCGGACACGAGACCCCAACACGCCCGCGGAGAGGCGCGNCGACAAG 358
QY 62 GGTGACCAAGTCGACCGAGTGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGGCAT 121
Db 359 CCGCCGCGCGACGAGCGCGCGCGCGCACGCCANCGACNCGCGCGCGCGCGAC 418
QY 122 GCCGAACACCGGCCAGGATCAGTCGAGCAGCGCGATGCGCGCTTCGACGGT 181
Db 419 GGCAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 478
QY 182 CGGCGCGAGCTCGATCTGTCGCCCGGCTCCACACAGGGGAACTGGTCGTCGCGAG 241
Db 479 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 538
QY 242 CGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 301
Db 539 AGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 595

```

```

QY 302 AGCGGAAACGGCGACGCGCTCAGATCCCTGTGTCGATCGCATCGGCTCAGTCCGCTGTCGCC 361
Db 596 CGCGAGGACCGCGCCACGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655
QY 362 CTTGCGCTTGGAGATAGCGTTTACGACGAGCGGCGACCAAGCGGCGCGCGCGCGCGCGG- 420
Db 656 ACGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715
QY 421 GCGGTTTACGCGATCCGCTCGATGACACGCGCTCTCGCGGGTTCGCGGGCGGTCGCGGATCC 480
Db 716 GCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775
QY 481 CGTACCGCGCGAGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520
Db 776 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 815

RESULT 17
CF876206 835 bp mRNA linear EST 31-OCT-2003
LOCUS trico40xn11.bl T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico40xn11, mRNA sequence.
ACCESSION CF876206
VERSION CF876206.1 GI:38130888
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
REFERENCE Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
AUTHORS 1. (bases 1 to 835)
Diener,S.E., Dankmeyer,L., Dunn-Coleman,N., Houfek,T.D.,
Mitchell,T.K., van Solingen,P., Teunissen,P.J.M., Ward,M. and
Dean,R.A.
TITLE Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
JOURNAL Unpublished (2003)
COMMENT Contact: Ralph A. Dean
Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: LT-F1 primer.
Location/Qualifiers
1. 835
/organism="Hypocrea jecorina"
/mol_type="mRNA"
/strain="QM6a"
/db_xref="taxon:51453"
/clone="trico40xn11"
/dev_stage="mycelia"
/clone_lib="T.reesei mycelial culture, Version 6 October
2003"
/note="Vector: pREP3y; Site 1: Not I/Sal I; Mycelial
culture grown from 24 hrs to 6 days with varying Carbon
and Nitrogen sources and concentrations."

ORIGIN
Query Match 11.3%; Score 58.8; DB 14; Length 835;
Best Local Similarity 46.7%; Pred. No. 7.9;
Matches 243; Conservative 0; Mismatches 273; Indels 4; Gaps 2;

QY 2 GTGCAGCAGCAACGCGGTGGTCCGCCGCCGCCATCACGAATGCCACTCCGCGG 61
Db 299 GGGCGCGGACACGAGACCCCAACACGCCCGCGGAGAGGCGCGNCGACAAG 358
QY 62 GGTGACCAAGTCGACCGAGTGGAGTTGAGCAGCCAGCTCATCGCCGCTGCGCGGCAT 121
Db 359 CCGCCGCGCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 418
QY 122 GCGGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 181

```



of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

```
ORIGIN
Query Match      11.3%; Score 58.6; DB 13; Length 1160;
Best Local Similarity 45.9%; Pred. No. 9;
Matches 162; Conservative 0; Mismatches 187; Indels 4; Gaps 1;

QY 129 CACGGGCCAGGATCAGTGCAGACGCGGATCGGCGCTTCGATCTCGAGCGTCCGCCGC 188
Db 1116 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057
QY 189 AGCTTCGATCTCTGTCGCGCGCGCTCCACACAGGCGGAACTGGCTCGGTGCGACGCGCAGC 248
Db 1056 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 997
QY 249 CCGAGCGGACAGCTCGTCAGGCTGCGTCTCAGGACCTCTCGCGGCTCGGAGCGGAA 308
Db 996 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 937
QY 309 ACGCG----CACGGCTCAGATCCCTGTCACTCGGCTCAGTCTCGGCTCGTCCCGCT 364
Db 936 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
QY 365 TGGCTCGGAGGATAGCGGTTACGACGAGCGGACACACGCGCGCGCGCGCGCGCGCGCG 424
Db 876 GCGCGCGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGG 817
QY 425 TTACAGCGCATCCGCTCAGTACACAGCGCTCGCGGGTTCGGGGTTCGGCGGAG 477
Db 816 GGGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGGNNNGGGGGG 764
```

```
RESULT 20
CNS006XXK
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence N7 end of BAC #
  BAC14N09 of RPII-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL066051.1 GI:4945019
VERSION
  GSS.
KEYWORDS
  Drosophila melanogaster (fruit fly)
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 935)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osoegawa and
  Aaron Mamooser in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPII-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
  Location/Qualifiers
    1. .935
    /organism="Drosophila melanogaster"
    /mol_type="genomic DNA"
```

/db\_xref="taxon:7227"  
/clone="BAC14N09"  
/clone\_lib="RPII-98"  
/note="end : 17"

```
ORIGIN
Query Match      11.2%; Score 58.4; DB 29; Length 935;
Best Local Similarity 28.8%; Pred. No. 9.3;
Matches 121; Conservative 107; Mismatches 191; Indels 1; Gaps 1;

QY 94 GCACGCTATCGCCGCTCGACGCTCGACGCTCGGCGCGGATCGATCTCGTCCGCGCTCC 213
Db 508 GCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 567
QY 154 CGGCGATCGGCGCTCGATCTCGACGCTCGGCGCGGCGCGCGCGCGCGCGCGCGCGCG 213
Db 568 GCGGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 627
QY 214 ACACAGGCGGAACTGGCTCGGTGCGACGCGGACGCCAGCCCGGAGACAGTCTGTCAGGC 273
Db 628 GCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 687
QY 274 TGGCGTCAAGCACTCTCTCGCGGTCGGGAGCGGAAACGCGCACGCTCAGATCCCTGTCA 333
Db 688 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 746
QY 334 CTCGATCGGCTCAGTCCGCTCGTCCCTTGGCTTGGGAGGATAGCGGTTCAAGCGA 393
Db 747 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
QY 394 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
Db 807 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 866
QY 454 TCGCGGCTCGGCGCGCTCGGCGGATCGTACCGCCCGGACCGCTCGGCGCGCGCGCG 513
Db 867 GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 926

RESULT 21
BZ569259/c
LOCUS
DEFINITION
  BZ569259 1015 bp DNA linear GSS 17-DEC-2002
  clone pac2-164_8319, genomic survey sequence.
ACCESSION
  BZ569259
VERSION
  BZ569259.1 GI:27203589
KEYWORDS
  Pseudomonas aeruginosa
  Pseudomonas aeruginosa
  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
  Pseudomonadaceae; Pseudomonas.
REFERENCE
  1 (bases 1 to 1015)
AUTHORS
  Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
  Burns, J.L., Kaul, R. and Olsen, M.V.
  Whole-Genome-Sequence variation among multiple isolates of
  Pseudomonas aeruginosa library
  J. Bacteriol. (2002) In press
  Contact: Chris K. Raymond
  Genome Center
  University of Washington
  Box 352145, Seattle, WA 98105-2145, USA
  Tel: 2062216954
  Fax: 2066857244
  Email: craymond@u.washington.edu
  Class: shotgun.
  Location/Qualifiers
    1. .1015
    /organism="Pseudomonas aeruginosa"
    /mol_type="genomic DNA"
    /strain="2-164"
    /db_xref="taxon:287"
    /clone="pac2-164_8319"
    /clone_lib="pac2-164"
```









```
RESULT 25
AG032118/c
LOCUS      1189 bp      DNA      linear      GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-006E05.R, genomic survey sequence.
ACCESSION  AG032118
VERSION     AG032118.1  GI:16558991
KEYWORDS   GSS.
SOURCE     Pan troglodytes (chimpanzee)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      BAC end sequences of Library PTB
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1189)
AUTHORS    Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
            Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail: chimpansegsc.riken.go.jp; URL: http://hgp.gsc.riken.go.jp/,
            Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB. This BAC end
            was generated during the R&D process and may have higher chance of
            clone-tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector      : pKS145
R.Site 1    : SacI
R.Site 2    : SacI
Location/Qualifiers
1. .1189
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-006E05.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN
Query Match      11.1%; Score 57.8; DB 29; Length 1189;
Best Local Similarity 44.2%; Pred. No. 12; Mismatches 283; Indels 0; Gaps 0;
Matches 224; Conservative 0; Mismatches 283; Indels 0; Gaps 0;
QY 10 CGAACGACCGGTGTCGCGCCCGCGCCCATCAGCAACTGCCACATCCGCGGGGTGACCA 69
DB 624 CGCCGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 565
QY 70 GTCGACAGGTGGAGTTAGCAGCCAGCTCATCGCGCCCTGCGCGCGCATGCCGAAC 129
DB 564 GTTGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 505
QY 130 ACGGCGCCAGGATCAGTGCAGCAGCGCGATCGCGCGCTCGATCTCGACGTCGCGCGCA 189
DB 504 CCACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 445
QY 190 GTCGATCTGTCGCGCGCGCTCCACACAGGGGAAAATGCGTCTCGGTGCGAGCGGACGC 249
DB 444 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 385
QY 250 CCAGCGCGGACAGCTCGTCAGGCTGGCGTACGAGACCTCTCGCGGTCGGGAGCGGAAA 309
DB 384 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 325
QY 310 CGCGCAGCGCTCAGATCCCTGTGTAFTCGATCGCTCAGTGCCTGCTGCTCCCTTGCC 369
```

```
DB 324 CGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 265
QY 370 TGGGAGGATAGCGGTTTACAGACAGCGGACACACGCGCGCGCGCGCGCGCGCGCGGTTCA 429
DB 264 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205
QY 430 CGATCCGCTCGATGACACGCGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCGCGGCTCG 499
DB 204 GGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 145
QY 490 CGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
DB 144 GCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 118

RESULT 26
AW731151
LOCUS      1387 bp      mRNA      linear      EST 08-MAR-2001
DEFINITION GA_Ea0010C13 Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA_Ea0010C13, mRNA sequence.
ACCESSION  AW731151
VERSION     AW731151.1  GI:7628809
KEYWORDS   EST.
SOURCE     Gossypium arboreum
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE  1 (bases 1 to 1387)
AUTHORS    Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
            Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
TITLE      An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL    Unpublished (2000)
COMMENT    Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            High quality sequence stop: 1387.
FEATURES
source
1. .1387
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ea0010C13"
/tissue_type="Fibers isolated from bolls harvested 7-10
            dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN
Query Match      11.0%; Score 57.4; DB 10; Length 1387;
Best Local Similarity 49.0%; Pred. No. 14;
Matches 177; Conservative 0; Mismatches 183; Indels 1; Gaps 1;
QY 156 GGGATCGCGCGCTCGATCTCGAGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 215
DB 851 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
QY 216 ACCAGGGGAAACTGGTTCGGTGGCAGCGGCGAGCCCGCGCGCGCGCGCGCGCGCGCG 275
DB 911 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 969
QY 276 GCGTCAGGACCTCTCGCGGTCGGGAGCGGAAACCGGACGCTCAGATCCCTGTCACT 335
DB 970 GCGCGCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1029
```





Class: shotgun.

## FEATURES

source  
1. .1610  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="2-164"  
/db\_xref="taxon:287"  
/clone="pacs2-164 8365"  
/clone\_lib="pacs2-164"  
/note="clinical isolate 2-164 Whole genomic shotgun library."

## ORIGIN

Query Match 10.9%; Score 56.8; DB 28; Length 1610;  
Best Local Similarity 46.1%; Pred. No. 18;  
Matches 224; Conservative 0; Mismatches 254; Indels 8; Gaps 2;  
QY 27 CCGCGCGCGCCATCAGCACTGCGACTCGCGGGGTGACACAGGTGACACAGTGGAG 86  
Db 1235 CG 1176  
QY 87 TTGAGCAGCCAGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 146  
Db 1175 CCTGCG 1116  
QY 147 TGCAGCAGCGCGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 206  
Db 1115 CG 1056  
QY 207 GCGTCCACACACAGCGGAAATGCTCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCG 266  
Db 1055 GCG 996  
QY 267 TCCAGGTGCGGTACACGACCTCTCGCGGTGCGGAGCGGAAACGCGACGGCTCAGATC 326  
Db 995 GCG 937  
QY 327 CTTGTCAGTCAGTCGCGTCACTAGTCGCGTCTGCTCCCTTGGCTGCGAGGATAGCGGTT 386  
Db 936 GCGC-----CGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 884  
QY 387 ACAGCAGCGCGCACACG 446  
Db 883 NGCG 824  
QY 447 CAGCGGTGCGGGGTGCGGGCGGTGCGGCGAGATCCGTACCGCGCGCGCGCGCGCGCG 506  
Db 823 CCG 764  
QY 507 GCGCGC 512  
Db 763 GCGCGC 758

## RESULT 31

LOCUS AJ442101/C  
DEFINITION AJ442101 dkf426 Gallus cDNA clone 19m6r1, mRNA sequence.  
ACCESSION AJ442101  
VERSION AJ442101.1 GI:20209322  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

## REFERENCE

1 (bases 1 to 483)  
Buerstedde, J.M.  
Gallus gallus bursal lymphocyte EST  
Unpublished (2002)  
Contact: Buerstedde JM  
Cellular Immunology  
Heinrich-Pette-Institute

Martinistr. 52, 20251 Hamburg, Germany

Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

## FEATURES

source  
1. 483  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="CB"  
/db\_xref="taxon:9031"  
/clone="19m6r1"  
/tissue\_type="Bursa of Fabricius"  
/cell\_type="Bursal lymphocyte"  
/dev\_stage="2-3 weeks old"  
/clone\_lib="dkfz426"

## ORIGIN

Query Match 10.9%; Score 56.6; DB 9; Length 483;  
Best Local Similarity 49.8%; Pred. No. 16;  
Matches 143; Conservative 0; Mismatches 144; Indels 0; Gaps 0;  
QY 32 GCGCGCCATCAGAACTGCGCTCGCGGGGTGACACAGGTGACACAGGTGGAGTTGAG 91  
Db 389 GCGGAGCAGCGCGGCHACAGGAGCCGACGACGACGACGACGACGACGACGACGAC 330  
QY 92 CAGCCAGCTCATGCGCGCGCTGCGCGCGCATGCCGGAACACCGGGCCAGGATCAGTGCGAG 151  
Db 329 CAGAGCAGCGCGCGGTACAGCGCGCGCCAGCCGAGCACAGGCGCGCCAGCAGAGCA 270  
QY 152 CAGCGCGATGCGCGCTCGATCTCGACGGTGGCGCGAGCTCGATCTCGTCCCGGTC 211  
Db 269 CAGCGCGCGCGCGCGCGCGCTGCGACAGCAGCGCGCGCGCGCGCGCGCGCG 210  
QY 212 CCACACACGCGGAACTGCGTGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 271  
Db 209 GCG 150  
QY 272 GCTGCGGTACGAGACCTCTCGCGGGTTCGGAGCGGAAACGCGCGCGG 318  
Db 149 GCGCTCGCGCGCTGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103

## RESULT 32

## LOCUS

## CG465713

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## Homo sapiens

## Homo sapiens

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## UNPUBLISHED (2003)

## CONTACT: KIM YS

## Human Genome Research Laboratory

## Korea Research Institute of Bioscience and Biotechnology

## 52 Boeun-dong, Yuseong-gu, Daejeon 305-333, Republic of Korea

## Tel: 82-42-879-8110

## Fax: 82-42-879-8119

Email: [yongsung@kribb.re.kr](mailto:yongsung@kribb.re.kr)

## This result was produced by the program for Functional Analysis of

## Human Genome funded from Ministry of Science and Technology,

## Republic of Korea. Clones are available from the Center for

## Functional Analysis of Human Genome. See URL:

<http://21cgenome.kribb.re.kr/> for details or contact:

## yongsung@kribb.re.kr.

## Seq primer: T7 primer

## Class: Noti site

## High quality sequence stop: 593.

Location/Qualifiers

1. 483

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="CB"

/db\_xref="taxon:9031"

/clone="19m6r1"

/tissue\_type="Bursa of Fabricius"

/cell\_type="Bursal lymphocyte"

/dev\_stage="2-3 weeks old"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

/clone\_lib="dkfz426"

```
FEATURES
  source
    Location/Qualifiers
      1..593
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /clone="KRIIB_2D_378"
        /sex="male"
        /tissue_type="blood"
        /dev_stage="adult"
        /clone_lib="pBS-NN Library"
        /note="Vector: pBluescriptKS(+); Site 1: NotI; Site 2:
        NotI; Genomic DNA was digested with NotI and the resulting
        fragments were ligated into ZAPII/NotI vector DNA. After
        the digestion with EcoRV, the mixture was run on 0.8% LMP
        agarose gel, eluted a gel portion over 22 kb and purified
        with beta-agarase I. The purified DNA was digested with
        NotI, run on 1% LMP agarose gel, and eluted with Gel
        Extraction kit the DNA fragments ranging 0.7 to 4 kb,
        which cover on first-dimensional gel comprising the
        central portion of the standard RIGS profile. The DNA
        ends was ligated into pBluescript KS(+) with NotI
        ends. The ligated solution was transformed into DH5a cell
        using electroporation machine."

ORIGIN
  Query Match      10.9%; Score 56.6; DB 29; Length 593;
  Best Local Similarity 48.6%; Pred. No. 16;
  Matches 245; Conservative 0; Mismatches 250; Indels 9; Gaps 3;

QY 3 TGCAGCAGACGACCGGTGTCGCCCGCGGCCCATACAGAACTGCCATCCGCGGG 62
Db 48 TTTCTACCGGGGGGAGCGGAGCCCTCGCGCTCGGGCTGCTGAGTGGGTAGC 107
QY 63 GTGACAGGTGACACAGGTGGAGTTGAGCAGCAGGTGATCGCGCTCGCGGATG 122
Db 108 GGCACCGGCCCCCGCGCGCTGGGGCAGCAGCGCAGCGCCGCGAGCCTCGAG 167
QY 123 CCGAAACACCGCGCCAGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTCGACGTC 182
Db 168 TCCAGCAGCTGGGGGNCACAGCGCGCTGCGCTGCGCTGCGCCGCGC-----C 222
QY 183 GCGCGACGTGATCTGTCGCCGCTCCACACAGCGGAAACTGCTCGGTGGCAGC 242
Db 223 TGCCGCTGCCGTGCGCGTGCACAGCCCCCGCGCGCAGCCCGCGCAGACGAGCAGC 282
QY 243 GGCACCCCGACCGGAGCAGTCTGTCAGGCTGCGTCAACGAGCTC--TCGCGGGTCGG 300
Db 283 CGGAGCCCGGACCGCGCGGAGGCTATCTGTGTGCAAGAGTTCTGCGCGCGCCT 342
QY 301 GAGCGAAACCGCAGCGCTCAGATCCCTGTGTCAGTCGATCGGTGCGGTGCTGTC 360
Db 343 GCGCGGCTCGCGAGACGAGTTCCACATCAGTGTCTATCAGGTGCGTGGCGCGGCTC 402
QY 361 CCTTTGGCTTGGAGGATAGCGGTTTACGAGCGGACCAACGCGGGCGGGCGGG 420
Db 403 TGGGGCGCGGGGACAGT--CAGGTGACCGGGCGGGGCGGGGCTCGGGGAGG 460
QY 421 GCGGTTTACGCGATCCGCTCGATGACAGCGGCTGCGGGGTGCGGGGTGCGGAGATC 480
Db 461 GCGCGTGGCGGGGGGTGCGGGGCGCGGGGAGCGGGGCGGGGCGGGGCGGGGAC 520
QY 481 CGTACCGCGCGACCGCTCGGCC 504
Db 521 CGGGACCGCAGCGCGGCAAGCC 544

RESULT 33
AG041117/c
LOCUS AG041117 976 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-018117.F, genomic survey sequence.
ACCESSION AG041117
VERSION AG041117.1 GI:16569842
KEYWORDS GSS.

SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 976)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
FEATURES
  source
    Location/Qualifiers
      1..976
        /organism="Pan troglodytes"
        /mol_type="genomic DNA"
        /db_xref="taxon:9598"
        /clone="PTB-018117.F"
        /sex="male"
        /cell_type="lymphoblast"
        /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
  Query Match      10.8%; Score 56.4; DB 29; Length 976;
  Best Local Similarity 45.3%; Pred. No. 19;
  Matches 224; Conservative 0; Mismatches 269; Indels 1; Gaps 1;

QY 17 CCGGTGGTCCCGCGCGCCCATCAGCAATGCGCATCTCCGCGGGGTGACAGGTGAC 76
Db 844 CCGCGNNGCCCGCGCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 785
QY 77 CAGGTGGGAGTTGAGCAGCAGCTCATCGCCGCTGCGCGCGCGCGCGCGCGCGCG 136
Db 784 GCGCCCGGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725
QY 137 -CAGGATCAGTCAGCAGCGGATGCGCGCTCGATCTCGACGGTTCGCGCGCGCG 195
Db 724 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
QY 196 TCTCTGCGCGCGCGTCCACACAGGGGAACTGGCTGCTGGCAGCGCGCGCGCGCG 255
Db 664 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
QY 256 GGGACAGCTCTGTCAGGCTGGCGTCAAGACCTCTCGGGGTTCGGGAGCGGAAAC 315
Db 604 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 545
QY 316 CCGCTCAGATCCCTGTGCTAGTCCATCGGCTAGTGGCTGCTGCTGCTGCTGCTG 375
Db 544 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485
QY 376 GATAGCGGTTTACGACGAGCGGACACACGCGCGGGGGGGCGGGCGCGGTTTTCAG 435
Db 484 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
QY 436 CGTCTGATGACAGCGGCTGCGGGGTTCGGGGTTCGGGGTTCGGGGTTCGGGGTTC 495
Db 424 GGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
```

D	b		676	GCCCCCGGCGCCGCCTGGCGGCCGGCGGCCGAGCACCACGCGCCCGCGGGCGG	735
Q	y		455	GCGGGTCGTGGGCGTTCGCGAGATCCGTACCCCCTCACCGCTTGCCAGCGCCGCG	514
D	b		736	GCGCGCCGCGGCG	795
Q	y		515	GG	516
D	b		796	CG	797
R	E	S	CNS015Y4	1203 bp DNA linear GSS 26-JUL-1999	
L	O	CUS	BACN15E10	Drosophila melanogaster genome survey sequence T7 end of BAC	
D	F	I	NITION	BACN15E10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.	
A	C	ESSION	ALU06054	GI:5619805	
V	E	RSION	ALU06054.1		
K	E	WORDS	GSS.		
S	O	RCE	Drosophila melanogaster (fruit fly)		
O	R	GANISM	Drosophila melanogaster		
R	E	FERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
A	T	HORS	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
T	I	TLE	Ephydroidea; Ephydriidae; Drosophila.		
J	O	URNAL	1 (bases 1 to 1203)		
C	O	MMENT	Genoscope.		
			Direct Submission		
			Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :		
			BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr		
			- Web : www.genoscope.cns.fr)		
			Determination of this BAC-end sequence was carried out as part of a		
			collaboration with the European Drosophila Genome Project (EDGP);		
			http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC		
			library (DrosBAC) was made by Alain Billaud at CEPH (Centre		
			d'Etude du Polymorphisme Humain) with funding provided by a MRC		
			project grant. The DNA was prepared from embryos by Alain Bucheton		
			and Genevieve Payan. It has been constructed in the vector		
			pBelobAC11.		
F	E	ATURES	Location/Qualifiers		
		s	1..1203		
		e	/organism="Drosophila melanogaster"		
			/mol_type="genomic DNA"		
			/db_xref="taxon:7227"		
			/clone="BACN15E10"		
			/clone_lib="DrosBAC"		
			/plasmid="pbeloBAC11"		
			/note="end : T7"		
O	R	I	GIN		
			Query Match	10.8%; Score 56; DB 29; Length 1203;	
			Best Local Similarity	32.2%; Pred.No. 23;	
			Matches	156; Conservative 84; Mismatches 245; Indels 0; Gaps 0;	
Q	y		29	CCGCGCCGCCCATCAAGAATTGCCACTCCGCGGGGTGACCAAGTCGACCAAGTTGGAGATT	88
D	b		675	CCGSGCSGSSCSSCSSSCSSCCSCCCSCSSGGGCGCGCGSGGGGGGGGGGGGGGG	734
Q	y		89	GAGCAGCCAGCTCATCGCGCGCTGCGCGCATGCGGAACAACGCGGCCAAGGATCAGTG	148
D	b		735	GGGGGSGGGCGCGCGCYCCCCCCCCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGG	794
Q	y		149	CAGCAGCGGATGCGCGCTTCGATCTCGAGGTCGCGCGCGTAGCTCGATCTCTGTCGCCCGG	208
D	b		795	GS GGCGCGSGGGGGGGCG	854
Q	y		209	CTCCCAACACAGGGGAACTGGTCTGTCGTCAGCGCAGCGCCCGCGCGCGCGCGCGCTC	268
D	b		855	CSGCCGGGGGGGGGNVCCSCCSCSGGGGGGSSCGCSSGGGGGGGGGGGGGGGGGGG	914
Q	y		269	CAGGCTGGGTACGAACTCTCGCGGTCGGGAGCGGAAAACGCGACGCTCAGATCCC	328
D	b		915	GGCGGGGWVGSGGGGVVVGGGVVMVGGSMGMGMVGMVGMVCGGGMGMSMCMGVGVVS	974



QY	329	TGTCAGTGCATCGGCTCAAGTCGGGTGCTGCCCTTGCCCTGGAGGATAGCGGTTAC	388
Db	975	GSSSGSGGGGGSGGGGGGYVGGCGVMMGGGGGGGGCCCCSSSGGGGGGMSGNMNGGGGGM	1034
QY	389	GACGACGCGCACCAACGCGGGGCGGGGGCGGFTTAGCCGATCCGCTCGCATGACCA	448
Db	1035	SSCGGSSSSGGGGGGGGGGGGCCCCSSSSSSSGSSSCCSCCCCSSSGCCCG	1094
QY	449	GGCGGTCGGGGGTCGGGGCGGTCGGCGAGATCCGTAACCGCCCGACCGCTCGCCAGCG	508
Db	1095	CCSCCCCSGSCCSCSSSGGGGGGGCGCCCGGGGGCCCCCGCGCCGCCGCCSCG	1154
QY	509	CGGCC 513	
Db	1155	SGGSC 1159	
RESULT 36			
BM912953/c			
LOCUS	BM912953	1449 bp	mRNA linear EST 12-MAR-2002
DEFINITION	AGENCOURT_6616214 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474939		
ACCESSION	S, mRNA sequence.		
VERSION	BM912953		
KEYWORDS	BM912953.1 GI:19363332		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 1449)		
JOURNAL	NTH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: DCTD/DTP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LHCMI988 row: p column: 12 High quality sequence stop: 274. Location/Qualifiers 1..1449 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:5474939" /tissue_type="amelanotic melanoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_41" /note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
FEATURES	source		
ORIGIN			
Query Match	10.8%;	Score 56;	DB 12; Length 1449;
Best Local Similarity	44.3%;	Pred. No. 24;	
Matches 214;	Conservative 0;	Mismatches 268;	Indels 1; Gaps 1;
QY	14	CGACCGGTGTCGCCCGCGCCCATACGAACTGCACCTCCGGCGGGTGACAGGTC	73
Db	935	CGCNGNNGCGTGC GCCGCCCGCCGCGCN CNNNNCNMGCCGCGGGGCGCGNGGC	876
QY	74	GACCAGGTGGGAGTTGAGCAGCCAGCTCATCGCGCTGCGCGGCGATGCCGAACACCG	133





Penniket, C., Roach, J.L. and Sarhan, F.  
Functional Genomics of Abiotic Stress in Wheat and Canola Crops  
Unpublished (2003)  
Contact: Wm L Crosby  
Bioinformatics  
University of Saskatchewan, Department of Computer Science  
1C101 Engineering Building, 57 Campus Drive, Saskatoon,  
Saskatchewan, S7N 5A9, Canada  
Tel: 306 966 1769  
Fax: 306 966 2033  
Email: fgas.ests@usask.ca  
This sequence is the direct result of the Base calling software  
phred (default parameters). It is the raw base calls. To aid in the  
identification of the high quality insert the software Lucy  
(default parameters) has been run on this sequence. Lucy identified  
the region [40,167].  
Plate: LAB002 row: K column: 11.  
Location/Qualifiers  
1. 1094  
/organism="Triticum aestivum"  
/mol\_type="rRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum PGAS: Library 4 Gate 8"  
/note="Organ: Crown and leaf; Vector: pCMV.SPOK16; Aerial  
parts (crown and leaf) of wheat cultivar Norstar during  
dehydration stress. 8 mRNA populations were combined  
before constructing the library. The first four come from  
removing plants from vermiculite (7 day old plants) and  
incubating them at 20C on the bench without water for 1,  
2, 3 and 4 days. The last four come from plants grown in  
soil in a growth chamber after watering is terminated.  
Four samplings were taken in a two week period; the first  
after wilting was observed and the last, two weeks later,  
consisted of live crown and stem tissue (leaf tissue was  
yellow and dead). First strand synthesis in this library  
was done in the presence of methylated dCTP thereby  
protecting from internal cleavage with NotI."

FEATURES

source  
1. 1094  
/organism="Triticum aestivum"  
/mol\_type="rRNA"  
/db\_xref="taxon:4565"  
/clone\_lib="Triticum aestivum PGAS: Library 4 Gate 8"  
/note="Organ: Crown and leaf; Vector: pCMV.SPOK16; Aerial  
parts (crown and leaf) of wheat cultivar Norstar during  
dehydration stress. 8 mRNA populations were combined  
before constructing the library. The first four come from  
removing plants from vermiculite (7 day old plants) and  
incubating them at 20C on the bench without water for 1,  
2, 3 and 4 days. The last four come from plants grown in  
soil in a growth chamber after watering is terminated.  
Four samplings were taken in a two week period; the first  
after wilting was observed and the last, two weeks later,  
consisted of live crown and stem tissue (leaf tissue was  
yellow and dead). First strand synthesis in this library  
was done in the presence of methylated dCTP thereby  
protecting from internal cleavage with NotI."

ORIGIN

Query Match 10.7%; Score 55.6; DB 14; Length 1094;  
Best Local Similarity 42.8%; Pred. No. 26;  
Matches 202; Conservative 0; Mismatches 266; Indels 4; Gaps 2;  
Qy 49 GCACCTCGCGGGGTGACAGGTGACAGGTGAGGTGAGCAGCAGCTCATCGCG 108  
Db 295 GCCTGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354  
Qy 109 CTGCGCGCGGATGCGGAAACACCGCGGCGCAGGATCACGTGACGACGCGCGCT 168  
Db 355 CCGCGCGGCG 414  
Qy 169 CGATCTCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 228  
Db 415 NCGNCCCG 474  
Qy 229 GCGTGGTGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 288  
Db 475 GGNNGNCCCG 534  
Qy 289 CTGCGCGGTGCGGAGCGGAAACCGCGACGCGCTCAGATCCCTGTCTAGTCTG 348  
Db 535 GCG 594  
Qy 349 TGCGGCTGCTGCC--CTTGGCTGGAGGATAGCGGTTACGACGAGCGGACACG 406  
Db 595 CCGCGNCCCG 654  
Qy 407 GCGCGCGCGCGCGCGCGGTTCAGCCGATCCGCTCGATGACACGCG--CTGCGG 464  
Db 655 CGCCCG 714  
Qy 465 GCGGCTGCGGAGATCGGTACCGCGGACCGCGCTCGGCGGCGCGCGCG 516

Db 715 CCG 766  
RESULT 40  
LOCUS CNS015Y4/c  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC  
BACN15E10 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION AL106054 GI:5619805  
VERSION AL106054.1  
KEYWORDS Drosophila melanogaster (fruit fly)  
SOURCE Drosophila melanogaster  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Phytophagae; Drosophilidae; Drosophila.  
1 (bases 1 to 1203)  
Genoscope.  
Direct Submission  
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the European Drosophila Genome Project (EDGP) -  
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC  
library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
d'Etude du Polymorphisme Humain) with funding provided by a MRC  
project grant. The DNA was prepared from embryos by Alain Bucheton  
and Genevieve Payan. It has been constructed in the vector  
pBelobAC11  
Location/Qualifiers  
1. 1203  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone\_lib="BACN15E10"  
/clone\_lib="DrosBAC"  
/plasmid="pBelobAC11"  
/notes="end : T7"  
ORIGIN  
Query Match 10.7%; Score 55.6; DB 29; Length 1203;  
Best Local Similarity 35.3%; Pred. No. 27;  
Matches 164; Conservative 81; Mismatches 215; Indels 5; Gaps 2;  
Qy 55 CCGCGCGGTTGACAGGTGACAGGTGAGGTGAGGTGAGCAGCTCATCGCGCTGG 114  
Db 1175 CCGCGGSGCSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 1116  
Qy 115 CCGCATGCGGAAACACCGGCGCAGGATCACGTGACGACGCGCGCTCGATCT 174  
Db 1115 CCGSGGSG 1056  
Qy 175 CGACGTTGCGCGCGAGCTGATCTGTCGCGCGCGCTCCACACGAGGGGAAATG 234  
Db 1055 CCG 996  
Qy 235 GTGCGACGCTC 291  
Db 995 CBCCCCCG 936  
Qy 292 GCGGTTGCG--GAGCGGAAACGCGCGAGCTGAGTCCCTGTGAGTCTGAGTCT 349  
Db 935 CCBBCCG 876  
Qy 350 GCGGTTGCG 409  
Db 875 SGGNCCCG 816  
Qy 410 CCGCGCGCGCGCGCGCGTTTTCAGCCGATCCGCTCGATGACAGCGGTCGCGG 469  
Db 815 CCG 756



158	GGCGAGCAGCGCGCGCGGAGACCGTATCTGTACCATGGCGATGGAGGGCTACGACCG	217
62	GGTGACCAAGGTCCGACCAAGGTGGGAGTTGAGCAGCCAGCTCATCGCGCCTTCGGCGCGGAT	121
218	CGAGTTCTGGCAGTTCCGGCAGCCAGCTCGCGCTCCAGACGGCCACCTTTCTCCGNNCGCT	277
122	GCGGAAAACACCGGGCCAGGATCACTGCAGCACGGCGATCGCGGCTCGATCTCGAC--G	179
278	CTCCNNNTTGGCGACTCCATCTGTGTGTCGCCCGCCGACGCCNNCCCCNNNGNCNCGCNN	337
180	GTCCGCGCCGAGTCCGATCTCTGTCGCCCGGTCCCAACACAGGGGAACTGGTCTCGTGGC	239
338	GCCNNCCCGGCCCCNNCNCNCCCGCCCGGGCGGGCCCCGCCNNGCCCGCCNCCG3NNC	397
240	AGCGGAGCCCCAGCGCGGGAAGCTCTGTCAGGCTGGCGTCAAGACCTCTCGCGGGTCG	299
398	GGCGGCGCGCCNCGGGCGCGCTCCNCGGGGCCCGCGCGCGCNCNCGCGCCGCC	457
300	GGAGCGGAAACGCGCACCGCTCAGATCCCTGTCAGTGCGCATCGGCTCAGTGCGGGTCTC	359
458	GCCNCGGGGCCCCNCCCGCGGNCNCGCCCGCGCCGCGCCNCCNCGNGGCTTCGCCGCG	517
360	CCCTTTGGCCTGGAGGATAGCGGTTCACGACGAGCGGCAACACGCGGCGCGGGCGCGG	419
518	GCCCCCGGCGCC-----NGGCCCCNCGCCNCGCGCCCGCGGGCCGNGCGGGGCGNNG	568
420	GGCGGTTACAGCGATCCGCTCGATACACAGCGGTCGCGGGTCGGGGCGGTCCGCGCAGAT	479
569	GCCGNCNCCNNGNNGCCCGCGGNCNCCCGCGCGCCGCGCNCCTTGGCGCCCGCGCGGGG	628
480	CGGTACCGCCCGGAACGCGCTCGGCCAGCG	508
629	CCCGCGCGCCCGNCGGGGGNCGCGGGG	657

RESULT	43
AG077581	:
LOCUS	: AG077581
DEFINITION	Pan troglodytes DNA, clone: PTB-072003.F, genomic survey sequence.
LINEAR	DNA linear GSS 03-NOV-2001
SIZE	1086 bp
CYCLES	PTB-072003.F

RECORDS	SOURCE	ORGANISM	GENUS
		Pan troglodytes (chimpanzee)	
		Pan troglodytes	
		Eukaryota; Eutezoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.	

REFERENCE	1
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	BAC end sequences of Library $\Phi$ T8

JOURNAL  
REFERENCE  
AUTHORS

**JOURNAL TITLE**  
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: [chimbese@gsc.riken.go.jp](mailto:chimbese@gsc.riken.go.jp), URL: <http://hgp.gsc.riken.go.jp/>, Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENTS: Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS  
Sequencing: -21M13

LIBRARY  
Vector : pKS145

```
R.Side 1 : sac1
R.Side 2 : sac1.
```

FEATURES	Location/Qualifiers
source	1..1086

```

:
:
: /organism="Pan troglodytes"
: /mol type="genomic DNA"

```

# Introduction

```

/db_xref="taxon:9598"
/_clone="PTB-072C03.F"
/_sex="male"
/_cell_type="lymphoblast"
/_clone_libs="PTB Chimpanzee Male BAC Library"

```

## ORIGIN

Query Match 10.7%; Score 55.4; DB 29; Length 1086;  
Best Local Similarity 37.7%; Pred. No. 28;  
Matches 193; Conservative 0; Mismatches 315; Indels 4;

	Qy	60	G G T G C A G C A G C G A A C G A C C G G T G G T C G C C C C G G C C G C C A C A C G A A C T G C C A C T C C G G G C	60
1	Db	491	G G C G C G C C C G N N N N N N N N N N N N N C G C G G C C G G G C C C C G C C C C C C N C N	550
	Qy	61	G G G T G A C C A G S T C G A C C A G S T G G A G T T G A G A C C A G C T C A T C G C C G C C T G C C C G G C A	120
	Db	551	N N N N C C N N N N C C C C C C N G C C N N N G G N G N G C C G G C C C G G C C C C C N N C N N N N N N	610
	Qy	121	T G C C G A A A C A C G G G C C A G G A T C A G T G C A G A C G S G C G A T G C G G C C T C G A T C T C G A C G G	180
	Db	611	N N N N N N N N N G G G C G G G G G G N N N N G N G N G N G G G G G G G G G C C C C C C C C C C C C N C C	670
	Qy	181	T C G C G C C A G A G T C G A T C T C T G C C G C C C G G C T C C C A C A C C A G G S G A A A T G G C T C G S T G G C A	240
	Db	671	G C C N N N C N N C C C G C G S G G C G S G G N G C C C C C C C G C G N N C C C C G C G G G G G G G G G	730
	Qy	241	G G G C A G C C C C A G C C G G A C A G T C T G T C C A G S T G T G G T C A C C G A C C T C T C G C G G G T C G G	300
	Db	731	G G G G G G G G G G C C C N C N C N G N N N G N N G N N G G G C G G C C C G S G G G N N G C C G G S G C C G G C	790
	Qy	301	G A G C G A A A C G C G C A C C G G C T C A G A T C C C T G T C A G T C G C A T C G G T C A G T G C C G T C G T C C	360
	Db	791	C G G C G N N N C C G C N N G G G C N N N N C C C N N G G N N C G G C C C C C C C G G G C C G S G G G N C N	850
	Qy	361	C C C T T G C C T G G A G A T A G C G T T C A C G A C A G C G G C A C C A C G G C G G G C G G S G G C G G G G	420
	Db	851	C C C C C C C C G G N C C G G G C C C - - - - C C C G G C N C G C G G C C C C C C G G C G G G G G G C G S C G C	906
	Qy	421	G G G T T C A G C C G A T C C G T C G A T G A C A G C G G C T G C G S G G T C G S G G G C G T C G G C G A G A T C	480
	Db	907	C C G C G C C C C N G S G G G C G G C N N C C C G B C G S G G C C N G G N N C C G C G S G G G C C C C C G C C	966
	Qy	481	C G T A C C G C C C G A C C G C C T C G C C A C A G C C C G C	512
	Db	967	N N C C C C C C C G G C C C C G C G N N G G C G G G G G G C	998

RESULT 44

AG072425/C  
LOCUS

DEFINITION  
ACCESSION

RECEIVED  
VERSION  
KEYWORDS

REVIEWED  
SOURCE  
ORGANISM

5445

## REFERENCE

TITLE	AUTHORS
1. The Role of the Teacher in the Classroom	John Doe
2. The Impact of Technology on Education	Jane Smith
3. The Importance of Parental Involvement	Michael Brown
4. The Challenges of Modern Education	Sarah White
5. The Future of Learning	David Green

FILE  
JOURNAL  
REFERENCE

## REFERENCE AUTHORS

TITLE

JOURNAL



TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: sheared ends.

FEATURES  
source  
Location/Qualifiers  
1..912  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA060B22"  
/clone\_lib="ZM\_0.7-1.5\_KB"  
/note="Vector: pBESK-; Site 1: HindII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

ORIGIN  
Query Match 10.6%; Score 55.2; DB 29; Length 912;  
Best Local Similarity 47.2%; Pred. No. 30;  
Matches 188; Conservative 0; Mismatches 188; Indels 0; Gaps 0;  
QY 82 GGGAGTTGAGCAGCCAGCTCATCGCGCTCGCGCATCGCGCATGCCGAACACCGGGCCAGGA 141  
Db 670 GGGGGTCCAGGAGCATGGCTCGCCATCTCGGCCACGAGGCTGGGCGAGTCCAGGTCCA 611  
QY 142 TCACGTGACGACGGGATCGCGCCTCGCATCTCGACGGTCGGCGCGCAGCTCGATCTCGT 201  
Db 610 GCTCGAATCTCTGTCGCTCCGAGCTCGACCAACCGGCGCGCGCGCCCGGACCTCGT 551  
QY 202 CGCCCCGCTCCCAACACGAGGGAAATGGGTCTCGGTGGCAGCGGACGCCAGCCCGGGACA 261  
Db 550 GGTGCGGTGCTAGTCTGTGGCACCGCGCTGCCCGCCCTGCTGGAGCAGCAGCAGCGCG 491  
QY 262 GCTCTGTCAGGTGGGGTTCACGACCTCTCGCGGGTGGGAGCGGAACCGCACGGCTC 321  
Db 490 GGTGCGAGCGACCGCGGTCTGCGGCTCGCGCCCGCGCGCGGACGTCGCGCGCGTGG 431  
QY 322 AGATCCCTGTCAGTCGCATCGGCTCAGTGCAGCGGTGCTCCCTTGGCTGGGAGGATAGC 381  
Db 430 AGTTCGCGGGCGCGCAGCCAGGCGACGAGCCCGGGAGTTGAGTCGCGCTCGCGGC 371  
QY 382 GGTTCACGACGAGCGACACACGGCGGGCGGGGGCGGGCGGTTCAGCGCATCCG 437  
Db 370 CCGCGAGGCGCAGCGCGGCGACGCTGTGGGCCACCGCGGCGCATCTCGCGCGACTCG 315

RESULT 48  
BM562099/c  
LOCUS  
DEFINITION  
5', mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
BM562099 1273 bp mRNA linear EST 20-FEB-2002  
AGENCOURT 6597561 NIH\_MGC\_41 Homo sapiens cdna clone IMAGE:5480748  
BM562099  
BM562099.1 GI:18807897  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1273)  
NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: DCTD/DTF  
cdna Library Preparation: Rubin Laboratory  
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.  
 Location/Qualifiers  
 1..633  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMBMA0595G13"  
 /clone\_lib="ZM 0.7 1.5 kb"  
 /note="Vector: pBC5K; Site1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN  
 Query Match 10.6%; Score 55.2; DB 29; Length 633;  
 Best Local Similarity 47.2%; Pred.No.28;  
 Matches 168; Conservative 0; Mismatches 188; Indels 0; Gaps 0;  
 QY 82 GGGAGTTGAGCAGCCAGCTCATCGCGCTGCGCCGCGCATGCCGAACACCCGGCCAGGA 141  
 Db 506 GCGGGTCCAGGACATGGGCTCCGCCATCTCGGCCCCAGAGGCTGGCGAGTCCAGGTCCA 447  
 QY 142 TCACGTGCAGCAGCGGATGCGGGCTCGATCTCGACGGTCGGCCCGCAGCTCGATCTCGT 201  
 Db 446 GCTCGAATCTCTGTCGCTCCGAGCTCGACACCGGCGCCGCCGCACTCGT 387  
 QY 202 CGCCCGGCTCCACACAGGGGAAAATGGGCTCGGTGGCAGCGGAGCCCGAGCCGGGACA 261  
 Db 386 GGTGCTGGTGTAGTGTGTGACCGGCGCTCCGCCCGCTCTGTGAGCAGCAGCAGCGCCG 327  
 QY 262 GCTCGTCCAGGCTGGGCTGCACGACCTCTCGCGGTCGGAGCGGAAACGCGACCGCTC 321  
 Db 326 GGTTCGAGCGCACCGCGTCTGCGGCTTCGGCCCGCGCGCGAGCGTCCGCGCGTGCG 267  
 QY 322 AGATCCCTGTTCAGTCGCATCGGCTCAGTGCCTGCTGCCCTTGGCTTGGAGGATAGC 381  
 Db 266 AGTTCGCGCGCGCGCAGCAGCGACGCGCCGGGAAGTTGAGTCGCGCTCGCGGC 207  
 QY 382 GGTTCAGCAGCAGCGGACACACGGCGGCGGGCGGGCGGCTTACCGCATCCG 437  
 Db 206 CCGCAGGGCGCAGCGCGCCACGCTGTGCGCCACCGCGCCATCTCGGCGCACTCG 151

RESULT 47  
 CG328665/c  
 LOCUS  
 DEFINITION  
 CG328665  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Zea mays  
 Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 912)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
 Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSEs: OGMJF117V  
 Contact: Cathy Whitelaw

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

Plate: LCM2004 row: b column: 13  
 High quality sequence stop: 206.  
 Location/Qualifiers  
 1. .1273  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGS:5480748"  
 /tissue\_type="amelanotic melanoma, cell line"  
 /lab\_host="NIH MGC 41"  
 /clone\_lib="NIH MGC 41"

features

#### FEATURES

source

1. .734  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST368j17"  
 /dev\_stage="36"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN59"  
 /note="Organ: limbs; Vector: pBluescript II KS(+); Site: 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaudo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

#### ORIGIN

Query Match 10.6%; Score 55.2; DB 12; Length 1273;  
 Best Local Similarity 50.0%; Pred. No. 31;  
 Matches 155; Conservative 0; Mismatches 154; Indels 1; Gaps 1;  
 202 CGCCCGGCTCCACACAGGGGAACTGGTCTGGTGGACGCGACGCCCGGACCA 261  
 |||||  
 Db 971 CMCCTCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912  
 262 GCTGTCAGGCTGGCTCAGGACCTCTCGCGGTGCGGAGCGGAAACGCGACGCTC 321  
 |||||  
 Db 911 GCGGGGCG 852  
 322 AGATCCCTGTGTCAGTCGATCGGTCAGTGGCGGTGCTCCCTTGGCTGGAGTAGC 381  
 |||||  
 Db 851 GCG 792  
 382 GTTTCAGACAGAGCGGACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 441  
 |||||  
 Db 791 CGGCG 733  
 442 ATGACACGACGCGCTCGCGGCGTGGCGGCGTGGCGAGATCCGTATCCGCGCGACCGCTCG 501  
 |||||  
 Db 732 CCCCCCGCGGCG 673  
 502 GCGAGCGCGG 511  
 |||||  
 Db 672 GGGCGCGCGG 663

RESULT 49  
 BU400981/c 734 bp mRNA linear EST 27-NOV-2002  
 LOCUS 603481269P1 CSEQCHN59 Gallus gallus cDNA clone CHEST368j17 5', mRNA  
 DEFINITION sequence.  
 BU400981  
 BU400981.1 GI:25770037  
 VERSION  
 EST.  
 KEYWORDS  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 734)  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 A Comprehensive Collection of Chicken cDNAs  
 Curr. Biol. 12 (22), 1965-1969 (2002)

REFERENCE  
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 TITLE A Comprehensive Collection of Chicken cDNAs  
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)  
 MEDLINE 22335534  
 PUBMED 12445392  
 COMMENT  
 Contact: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology

(UWIST)

PO Box 88, Manchester, M60 1QD, UK  
 Tel: 0161208930  
 Fax: 01612360409  
 Email: Simon.Hubbard@umist.ac.uk.  
 Location/Qualifiers

#### FEATURES

source

1. .734  
 /organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST368j17"  
 /dev\_stage="36"  
 /lab\_host="DH10B"  
 /clone\_lib="CSEQCHN59"  
 /note="Organ: limbs; Vector: pBluescript II KS(+); Site: 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaudo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

#### ORIGIN

Query Match 10.6%; Score 55; DB 13; Length 734;  
 Best Local Similarity 49.5%; Pred. No. 31;  
 Matches 142; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
 QY 32 GCGCGCCATCATCGAACTCCCACTCCGCGGGGTGACCAAGTTCAGGAGTTGAG 91  
 |||||  
 Db 377 GCGGAGCAGCGCGCACAGGAGCCAGGACGACGAGCGAATTCGCGGCGCGAG 318  
 92 CAGCCAGCTCATCGCGCGCTGGCGGATCGGAAACACCGCGGACGATCAGTCGAG 151  
 |||||  
 Db 317 CAGAAGCAGCGCGGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258  
 152 CACGCGGATGCGCGCTCGATCTCGACGTCGCGCGCGCGCGCGCGCGCGCGCG 211  
 |||||  
 Db 257 CAGCG 198  
 212 CCACACAGGGGAACTGGTCTGGTGGACGCGGAGCCCGCGCGGACAGCTGCTCAG 271  
 |||||  
 Db 197 GCGCGCCACGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 138  
 272 GCTGCGTCACGAGACTCTCGCGGTGCGGAGCGGAAACGCGCACCG 318  
 |||||  
 Db 137 GCGCTCCGCGCTGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91

#### RESULT 50

AG058001/c

LOCUS

DEFINITION

AG058001

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

2 (bases 1 to 1285)

AG058001 1285 bp DNA linear GSS 02-NOV-2001  
 Pan troglodytes DNA, clone: PTB-04M04.F, genomic survey sequence.

AG058001.1 GI:16595462

GSS.

Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

1

Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,

Totoki, Y., Watanabe, H. and Sakaki, Y.

BAC end sequences of Library PTB

Unpublished

2 (bases 1 to 1285)



**AUTHORS** Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,  
Totoki,Y., Watanabe,H. and Sakaki,Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehico-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,  
Tel:81-45-503-9111, Fax:81-45-503-9170)  
**COMMENT** Clones are derived from the chimpanzee BAC library PTB This BAC end  
was generated during the R&D process and may have higher chance of  
clone tracking errors.  
**PRIMERS**  
Sequencing: -21M13  
**LIBRARY**  
Vector : PKS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
Location/Qualifiers  
1. .1285  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
/clone="PTB-044M04.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_lib="PTB Chimpanzee Male BAC Library"

**FEATURES**  
source

**ORIGIN**

Query Match 10.6%; Score 55; DB 29; Length 1285;  
Best Local Similarity 46.7%; Pred. No. 34;  
Matches 195; Conservative 0; Mismatches 219; Indels 4; Gaps 1;  
QY 102 ATCGCGGCTGCGCGGATGCCGAAACACCGGGCCAGGATCAGTGCACGACGCGGATG 161  
Db |||||  
553 ACCGGCGCGGGCGCGGGCGCGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 494  
QY 162 CGCGCTCTGATCTCGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 221  
Db |||||  
493 CG 434  
QY 222 GGAAGTGGCTCGTGGGAGCGGCGAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 281  
Db |||||  
433 CCGNCG 374  
QY 282 CGGACTCTTCGCGGGTCGCGGAGCGGAAACCGCGCACGCGTCAAGATCCCTGTCAGTCGCATC 341  
Db |||||  
373 CCCCTCG 318  
QY 342 GGCTCAGTCCGCTGTCCTCCCTTGGCTTGGAGAGATAGCGGTTTCAGACGAGCGGACC 401  
Db |||||  
317 GCCCCCCCGNCG 258  
QY 402 ACGGCGGGGCGGGGCGGGGCGGTTTCAGCGGATCCGCTCGATGACACGCGGCTGCGGGGT 461  
Db |||||  
257 GCGGCG 198  
QY 462 CGGGGCGGTTCGCGGAGATCCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519  
Db |||||  
197 CGGGGCG 140

Search completed: June 27, 2004, 20:50:49  
Job time : 2115.57 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 27, 2004, 18:07:09 ; Search time 2364.22 Seconds  
(without alignments)  
9166.472 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_1\_500  
Perfect score: 500  
Sequence: 1 ggtaaccgacgggtccggg.....tgcacgtcgccggaacct 500

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : GenBank

- 1: gb\_ba.\*
- 2: gb\_hg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pt.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	500	100.0	37116	6	AX574199	Sequence
C 2	500	100.0	109519	6	AX195929	Sequence
C 3	369	73.8	1251	6	AX574152	Sequence
C 4	366	73.2	1248	6	AX195930	Sequence
C 5	298.4	59.7	10035	6	AX574202	Sequence
C 6	296.4	59.7	11115	6	AX205014	Sequence
C 7	256.6	51.3	1251	6	AX574150	Sequence
C 8	216	43.2	37541	1	AOPCZA361	Sequence
C 9	208.8	41.8	56917	1	AME16952	Sequence
C 10	163.4	32.7	18977	1	SAR7932	Sequence
C 11	149.8	30.0	1401	6	AR088311	Sequence
C 12	149.8	30.0	1401	6	AR144738	Sequence
C 13	149.8	30.0	2221	1	AF048833	Streptomy
C 14	135	27.0	16020	6	BD251846	Gene clus
C 15	135	27.0	16065	1	AF187532	Streptomy
C 16	131	26.2	12152	6	AX196110	Sequence
C 17	124	24.8	603	6	AX574140	Sequence
C 18	118.2	23.6	15240	6	AX574198	Sequence
C 19	116.2	23.2	31310	1	AX131229	Streptomy
C 20	112.8	22.6	1236	6	AX195978	Sequence
C 21	112.8	22.6	1281	6	AX574042	Sequence
C 22	108.2	21.6	4254	1	SERVAB10	Sequence
C 23	105.6	21.1	315000	1	RME603644	Sequence
C 24	104.2	20.8	5908	1	AF147704	Sequence
C 25	98.2	19.6	12862	1	AE008206	Agrobacte
C 26	98.2	19.6	12952	1	AE009408	Streptomy
C 27	97.2	19.4	291000	1	SCO93105	Streptomy
C 28	97	19.4	85163	1	AY048670	Streptomy
C 29	95	19.0	25617	1	AF170880	Sequence
C 30	95	19.0	25617	6	AX707121	Sequence
C 31	94	18.8	31248	6	AX707115	Streptomy
C 32	94	18.8	35359	1	AF235050	Streptomy
C 33	94	18.8	35359	6	AX707116	Sequence
C 34	92.4	18.5	1290	6	AX574040	Sequence
C 35	92.4	18.5	14252	6	AX204973	Sequence
C 36	92.4	18.5	48221	6	AX574200	Sequence
C 37	90.8	18.2	303100	1	RME603646	Rhizobium
C 38	90.6	18.1	45055	6	AX574197	Sequence
C 39	87.8	17.6	1224	6	AX574038	Sequence
C 40	86.2	17.2	210614	1	AB088224	Streptomy
C 41	81	16.2	50000	6	AX535264	Sequence
C 42	81	16.2	50000	6	AX535266	Sequence
C 43	81	16.2	59816	1	AF333038	Streptomy
C 44	80.8	16.2	1224	6	AX535308	Sequence
C 45	78	15.6	11659	1	AE009716	Brucella
C 46	78	15.6	11852	1	AE014542	Brucella
C 47	78	15.6	36502	1	AF076290	Brucella
C 48	78	15.6	42391	1	AF329398	Streptomy
C 49	78	15.6	42391	6	AX707118	Sequence
C 50	64	12.8	295150	1	SCO939126	Streptomy

ALIGNMENTS

RESULT 1  
AX574199/c  
LOCUS AX574199 37116 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 279 from Patent WO02079505.  
ACCESSION AX574199  
VERSION AX574199.1 GI:27551678  
KEYWORDS Micromonospora carbonacea  
SOURCE Micromonospora carbonacea  
ORGANISM Bacteria; Actinobacteria; Actinomycetales;

Micromonosporineae; Micromonosporaceae; Micromonospora.

## REFERENCE

1 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
Patent: WO 02079505-A 229 10-OCT-2002

## JOURNAL

Ecopia Biosciences Inc. (CA)

## FEATURES

Location/Qualifiers

1..37116

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

/note="var. africana"

## ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 37116;  
Best Local Similarity 100.0%; Pred. No. 2.1e-56; Indels 0; Gaps 0;  
Matches 500; Conservative 0; Mismatches 0;

Qy 1 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATACGGCAGAGAAACACCCCGGT 60  
Db 30486 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATACGGCAGAGAAACACCCCGGT 30427

Qy 61 AGTCCGGGTACAGCGTGGCGGGAAGCGTACGCGCTTCGACGCTCAGCGGGGGCGG 120  
Db 30426 AGTCCGGGTACAGCGTGGCGGGAAGCGTACGCGCTTCGACGCTCAGCGGGGGCGG 30367

Qy 121 ACACCGCGGCTCAGTCTGTCAGTGTACGCGGGGAGCGTACAGATCACTGTCCGCCA 180  
Db 30366 ACACCGCGGCTCAGTCTGTCAGTGTACGCGGGGAGCGTACAGATCACTGTCCGCCA 30307

Qy 181 GCCCGGCGGAACCTCCTGCTTCGCTCAGTGTACGCGCTTCGACGCTCAGCGGGGAGAGC 240  
Db 30306 GCCCGGCGGAACCTCCTGCTTCGCTCAGTGTACGCGCTTCGACGCTCAGCGGGGAGAGC 30247

Qy 241 AGCGGTAGTCCACCGCTCGCGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 300  
Db 30246 AGCGGTAGTCCACCGCTCGCGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 30187

Qy 301 CCGGGGTGAGCGGGGCGCTGCTTGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
Db 30186 CCGGGGTGAGCGGGGCGCTGCTTGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 30127

Qy 361 GGACCGATGCGCGAGAGTTCGTCAGGTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 420  
Db 30126 GGACCGATGCGCGAGAGTTCGTCAGGTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 30067

Qy 421 ACCGCTTGCCTCGCGCTTACGAGTTACGAGGCGAGCAGTTCGTCGCGATGCC 480  
Db 30066 ACCGCTTGCCTCGCGCTTACGAGTTACGAGGCGAGCAGTTCGTCGCGATGCC 30007

Qy 481 TCGACGTGCGCGGGAACCT 500

Db 30006 TCGACGTGCGCGGGAACCT 29987

## RESULT 2

AX195929

LOCUS

Sequence 1 from Patent WO0151639.

AX195929

ACCESSION

AX195929.1 GI:15386161

VERSION

KEYWORDS

SOURCE

ORGANISM

Micromonospora carbonacea

Bacteria; Actinobacteria; Micromonosporaceae; Micromonospora.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Location/Qualifiers

1

Hosted, T.J., Horan, A.C. and Wang, T.X.

Everminomycin biosynthetic genes

Patent: WO 0151639-A 119 JUL-2001;

Schering Corporation (US)

source

1..109519

/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 100.0%; Score 500; DB 6; Length 109519;  
Best Local Similarity 100.0%; Pred. No. 1.6e-56; Indels 0; Gaps 0;  
Matches 500; Conservative 0; Mismatches 0;

Qy 1 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATACGGCAGAGAAACACCCCGGT 60  
Db 1 GGTACCCGACCGTGTCCCGAACAACAGATCGAGATACGGCAGAGAAACACCCCGGT 60

Qy 61 AGTCCGGGTACAGCGTGGCGGGAAGCGTACGCGCTTCGACGCTCAGCGGGGGCGG 120  
Db 61 AGTCCGGGTACAGCGTGGCGGGAAGCGTACGCGCTTCGACGCTCAGCGGGGGCGG 120

Qy 121 ACACCGCGGCTCAGTCTGTCAGTGTACGCGGGGAGCGTACAGATCACTGTCCGCCA 180  
Db 121 ACACCGCGGCTCAGTCTGTCAGTGTACGCGGGGAGCGTACAGATCACTGTCCGCCA 180

Qy 181 GCCCGGCGGAACCTCCTGCTTCGCTCAGTGTACGCGCTTCGACGCTCAGCGGGGAGAGC 240  
Db 181 GCCCGGCGGAACCTCCTGCTTCGCTCAGTGTACGCGCTTCGACGCTCAGCGGGGAGAGC 240

Qy 241 AGCGGTAGTCCACCGCTCGCGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 300  
Db 241 AGCGGTAGTCCACCGCTCGCGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCT 300

Qy 301 CCGGGGTGAGCGGGGCGCTGCTTGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
Db 301 CCGGGGTGAGCGGGGCGCTGCTTGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360

Qy 361 GGACCGATGCGCGAGAGTTCGTCAGGTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 420  
Db 361 GGACCGATGCGCGAGAGTTCGTCAGGTCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCT 420

Qy 421 ACCGCTTGCCTCGCGCTTACGAGTTACGAGGCGAGCAGTTCGTCGCGATGCC 480  
Db 421 ACCGCTTGCCTCGCGCTTACGAGTTACGAGGCGAGCAGTTCGTCGCGATGCC 480

Qy 481 TCGACGTGCGCGGGAACCT 500

Db 481 TCGACGTGCGCGGGAACCT 500

## RESULT 3

AX574152/c

LOCUS

Sequence 232 from Patent WO02079505.

AX574152

ACCESSION

AX574152.1 GI:27551653

VERSION

KEYWORDS

SOURCE

ORGANISM

Micromonospora carbonacea

Bacteria; Actinobacteria; Micromonosporaceae; Micromonospora.

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Location/Qualifiers

1..1251

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

/note="var. africana"

## ORIGIN

Query Match 73.8%; Score 369; DB 6; Length 1251;

Best Local Similarity 100.0%; Pred. No. 5.2e-39; Mismatches 0; Indels 0; Gaps 0;

132 TCAGCTCGTACAGCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAA 191

1251 TCAGCTCGTACAGCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAA 1192

192 CTCCTGCTTCCGCAATGATCTCGTCGGGTGTTCCAGGCGAGAGCAGCGGTAGTC 251

1191 CTCCTGCTTCCGCAATGATCTCGTCGGGTGTTCCAGGCGAGAGCAGCGGTAGTC 1132

252 CACGGCTCGGGGCTGAACCGCTCCGGGTGCGCACCGGATGTCGTCGGGGGTGAG 311

1131 CACGGCTCGGGGCTGAACCGCTCCGGGTGCGCACCGGATGTCGTCGGGGGTGAG 1072

312 CCGGCTCTGTCGGCGCGTCTGTCGTCGACACCGGATGTCGTCGGGGGTGAG 371

1071 CCGGCTCTGTCGGCGCGTCTGTCGTCGACACCGGATGTCGTCGGGGGTGAG 1012

372 GCAGAGTTCGTCAGCGTGGCGTCTTCGCGCTCGCGCGTACGCCACACCGCTTGC 431

1011 GCAGAGTTCGTCAGCGTGGCGTCTTCGCGCTCGCGCGTACGCCACACCGCTTGC 952

432 CTCGGCTTACGAGTTCAGAGGCGGAGCGAGTTCGTCGGATGCCCTCGACGTCGGC 491

951 CTCGGCTTACGAGTTCAGAGGCGGAGCGAGTTCGTCGGATGCCCTCGACGTCGGC 892

492 GCGGAACCT 500

891 GCGGAACCT 883

RESULT 4

AX195930/c

LOCUS AX195930 1248 bp DNA linear PAT 28-AUG-2001

DEFINITION Sequence 2 from Patent WO0151639.

ACCESSION AX195930

VERSION AX195930.1 GI:15386162

KEYWORDS

SOURCE

ORGANISM

Micromonospora carbonacea

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Micromonosporinae; Micromonosporaceae; Micromonospora.

REFERENCE

1

AUTHORS Hosted, T.J., Horan, A.C. and Wang, T.X.

TITLE Evernimycin biosynthetic genes

JOURNAL Patent: WO 0151639-A 2 19-JUL-2001;

Schering Corporation (US)

FEATURES

source

1..1248

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

1..>1248

/note="unassigned protein product; evdA"

/codon\_start=1

/transl\_table=1

/protein\_id="CAC59991.1"

/db\_xref="GI:15386163"

/translation="MTSDAGTTPRGTACAGGTTAQFLDLGRPLSDRLFTDQLAD

EFFFLHAGVAGCTVMOLMEVEPREEMFHODYFHSYSSVMQKHPADTARHLLHE

ATGDPFVEIGNDYVRLVHAGVRLHGFEPGADVARGHGVVTFVPEEST

ARRVETGPDVLPFRANTICHPYLDLILGPDVLPVFPVFPVLPVLPVLPVLPV

PDQYDEHFFFSARSQVLAARFGPALVVERLPVHGGVVRVTPARAGARQARVA

ALIGEDARGVADSRDLRFADVEGIRLTLALNLNLSKAEKRVAYGATKASATV

NFGCGDPLVSVQCDRTTFAKQRLTPGTHIPVTPDPAVDVALLFAWNHAEIM

AKQEFRRAGGQWILYVPRVHTS"

## ORIGIN

Query Match 73.2%; Score 366; DB 6; Length 1248;

Best Local Similarity 100.0%; Pred. No. 1.3e-38; Mismatches 0; Indels 0; Gaps 0;

Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 GCTCGTACGTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAACTC 194

Db 1248 GCTCGTACGTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAACTC 1189

QY 195 CTGCTCTTCCCAATGATCTCGTCGGGTGTTCCAGGCGAGAGCAGCGGTAGTCAC 254

Db 1188 CTGCTCTTCCCAATGATCTCGTCGGGTGTTCCAGGCGAGAGCAGCGGTAGTCAC 1129

QY 255 CGCTCGGGCGTGAACCGGTCCGGGTGCGCACCGGATGTCGTCGGGGGTGAGCGG 314

Db 1128 CGCTCGGGCGTGAACCGGTCCGGGTGCGCACCGGATGTCGTCGGGGGTGAGCGG 1069

QY 315 GCTCTGCTTGGCCCGCTGTCGTCGACACCGGAGACCGAGTTCGAGTCCGCGCA 374

Db 1068 GCTCTGCTTGGCCCGCTGTCGTCGACACCGGAGACCGAGTTCGAGTCCGCGCA 1009

QY 375 GAAGTTCGTCAGGTGGCGCTCTTCGCGCTCGCGCGTACGCCACACCGCTTGCCTC 434

Db 1008 GAAGTTCGTCAGGTGGCGCTCTTCGCGCTCGCGCGTACGCCACACCGCTTGCCTC 949

QY 435 GGCCTTGAGCGAGTTCAGCAGGCGGAGCAGGTTCGTCGGATGCCCTCGAGTCCGGCGC 494

Db 948 GGCCTTGAGCGAGTTCAGCAGGCGGAGCAGGTTCGTCGGATGCCCTCGAGTCCGGCGC 889

QY 495 GAACT 500

Db 888 GAACT 883

RESULT 5

AX574202/c

LOCUS AX574202 10035 bp DNA linear PAT 07-JAN-2003

DEFINITION Sequence 282 from Patent WO02079505.

ACCESSION AX574202

VERSION AX574202.1 GI:27551681

KEYWORDS

SOURCE

ORGANISM

Micromonospora carbonacea

Micromonospora carbonacea

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Micromonosporinae; Micromonosporaceae; Micromonospora.

REFERENCE

1

AUTHORS Farnet, C.M., Zazopoulos, E. and Staffa, A.

TITLE Compositions and methods for identifying and distinguishing

JOURNAL orthomycin biosynthetic loci

Patent: WO 02079505-A 282 10-OCT-2002;

Ecopia Biosciences Inc. (CA)

FEATURES

source

1..10035

/organism="Micromonospora carbonacea"

/mol\_type="unassigned DNA"

/db\_xref="taxon:47853"

## ORIGIN

Query Match 59.7%; Score 298.4; DB 6; Length 10035;

Best Local Similarity 76.1%; Pred. No. 4.1e-30; Mismatches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;

QY 1 GGTACCCGACCGTGTCCCGGAACACGAGTTCGAGATACGGCGAGAGGAAACACCCCGGT 60

Db 3742 GATATCCGAGCGTCTCGTGAAGACCGAGTTCAGGTACGGGACACGAGACACCCCGG 3683

QY 61 AGTCCGGGTAGCGGTGGCGGAAAGCGGTACCGCGCTTCGACGGTACCGGGCGGCGG 120

Db 3682 GGTCCGGAAGACCCCGGGGTGAAGACCGTACCGCGCTTCGACGGGAGTTGCTGTA 3623

QY 121 ACACC---GGGGGTTCAGTTCGTCAGGTACGGGGGAGCGTACAGGATCCACTGTCC 176

Db 3622 CTGACAGGCGCCCACTCAATCCCGACGTGAACGTGCGGAACGTACAGATCCAGGCC 3563

QY 177 GCCAGCCCGCGGAACTCTCTGCTCTTCCGCATGATCTCGTGGGTGTTCCAGCGGAA 236

Db 3562 GCCGGCTGCGGAAAGCGCTGCTCTTCCGCATGATCTCGTGGGTGTTCCAGCGGAA 3503

QY 237 GAGCAGCGGTAGTCCACCGGTGGGCGTGAACGGCGTCCGGGTGCGCCACCGGATGTG 296  
Db 3502 GAGCAGCGGTAGTCCCGCGGCTGGTGGGAATCTGTCGCGCGCATGAACCGGATGTG 3443  
QY 297 CGTCCGGGGGTGAGCGCGGCTGCTGGCCGGGTGCTGTCGACACACCGAGGACGAG 356  
Db 3442 CGTCCCGGGGTGAGCGCGGCTGCTGGCCGGGTGCTGTCGACACACCGAGGACGAG 3383  
QY 357 GTCGCGGACCGATGCGCGAGAGTTCGTACCGTGGCGCTCTTCGCGCGTCCGGCGGTACGC 416  
Db 3382 GTCGCGGACCGATGCGCGAGAGTTCGTACCGTGGCGCTCTTCGCGCGTCCGGCGGTAGGC 3323  
QY 417 CACACCGCGCTTGGCTCGGCTTGGAGGAGTTCAGCGGCGGAGGAGTTCGTCGCGAT 476  
Db 3322 CACACCGCGCTTGGCTCGGCTTGGAGGAGTTCAGCGGCGGAGGAGTTCGTCGCGAT 3263  
QY 477 GCGCTCGACGTCGGCGGGAAC 498  
Db 3262 CCGCGCGACCTGGCGAGCGAAC 3241

RESULT 6  
AX205014/c 11115 bp DNA linear PAT 30-AUG-2001  
LOCUS AX205014 Sequence 49 from Patent WO0155180.  
DEFINITION AX205014  
ACCESSION AX205014  
VERSION AX205014.1 GI:15394258  
KEYWORDS  
SOURCE  
ORGANISM  
Micromonospora carbonacea  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE  
AUTHORS Staffa, A., Zazopoulos, E., Mercure, S. and Nowacki, P.  
TITLE Genetic locus for evernimycin biosynthesis  
JOURNAL Patent: WO 02079505-A 230 10-OCT-2002

FEATURES  
Location/Qualifiers  
1..11115  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

misc\_feature  
/note="ORF 41 (positive strandedness) incomplete;  
C-terminus only"  
1213..2331  
/note="ORF 42 (positive strandedness)"  
2364..3611  
/note="ORF 43 (positive strandedness)"  
3623..4243  
/note="ORF 44 (positive strandedness)"  
4149..5177  
/note="ORF 45 (positive strandedness)"  
5177..6094  
/note="ORF 46 (negative strandedness)"  
6271..7824  
/note="ORF 47 (negative strandedness)"  
7903..8760  
/note="ORF 48 (negative strandedness)"  
8781..9800  
/note="ORF 49 (negative strandedness)"

## ORIGIN

Query Match 59.7%; Score 298.4; DB 6; Length 11115;  
Best Local Similarity 76.1%; Pred. No. 4e-30;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
QY 1 GGTACCGCGGTGTCGCGGACACAGAGTTCGAGATACGGGAGAGGAGAACCCCGCGT 60  
Db 3749 GATATCCGAGCGTCTCGGTGAGACCGAGTTCAGGTACGGGACACGAGAACCCCGCG 3690  
QY 61 AGTCCGGGTAGACGCTGGCGCGAGAGCGGTACGCGCTTCGACGCTTCAGCGGCGGCGG 120

Db 3689 GGTCCGGGAAGACCCCGCGGGGTGAAGAGCGTACGCCCTTCGACGGCGAGTTCCGTTGCTA 3630  
QY 121 ACACC-----GGCGGGTCAAGTCTGTCAGCTGTACGCGGGGAGCTACAGATCCACTGTCC 176  
Db 3629 CTGTACAGCCCACTCAATCCGCACTGTAACGTGCGGAAGCTACAGATCCAGGCCCC 3570  
QY 177 GCGAGCCCGCGGAACTCTCTGCTCTTCCCAATGATCTGTCGGCGTGTTCAGGCGGAA 236  
Db 3569 GCGGCGCTCCCGGAAAGCGCTCTCTTCCCATGATCTGTCGGCGTGTTCAGGCGGAA 3510  
QY 237 GAGCAGCGGTAGTCCACCGGTGGGCGTGAACCGCTCCGGGTGCGCACCGGATGTG 296  
Db 3509 GAGCAGCGGTAGTCCGCGGCTCGTCCGGAACCTGCTCCGCGCATGAACCGGATGTG 3450  
QY 297 CGTCCGGGGGTGAGCGGCGCTCTGTCGCGGCGTCTGTCGACACCGAGGACGAG 356  
Db 3449 CGTCCCGGGGTGAGCGGCGCTCTGTCGCGGCGTCTGTCGACACCGGAGCACGAG 3390  
QY 357 GTCGCGACCGATGCGCGAGAGTTCGTACCGTGGCGGCTCTTCGCGCTCCGCGCTACGC 416  
Db 3389 GTCGCGCGGATGCGCGAGAGTTCGCAACGCTCGGCTCTTGGCGTCCGCGCTAGGC 3330  
QY 417 CACACCGCGCTTCCCTCGGCTTCGAGCGAGTTCAGCGGGGAGCAGGTCGTCGCGAT 476  
Db 3329 CACACCGCGCTTCCCTCGGCTTCGAGCGAGTTCGAGCGGAGCGCGCGAGTTCGTCGCGAT 3270  
QY 477 GCGCTCGACGTCGGCGGGAAC 498  
Db 3269 CCGCGCGACCTGGCGAGCGAAC 3248

## RESULT 7

AX574150/c 1251 bp DNA linear PAT 07-JAN-2003  
LOCUS AX574150 Sequence 230 from Patent WO02079505.  
DEFINITION AX574150  
ACCESSION AX574150  
VERSION AX574150.1 GI:27551652  
KEYWORDS  
SOURCE  
ORGANISM  
Micromonospora carbonacea  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE  
AUTHORS Farnet, C.M., Zazopoulos, E. and Staffa, A.  
TITLE Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci

JOURNAL Patent: WO 02079505-A 230 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)

FEATURES  
Source  
1..1251  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

## ORIGIN

Query Match 51.3%; Score 256.6; DB 6; Length 1251;  
Best Local Similarity 81.2%; Pred. No. 1.9e-24;  
Matches 298; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 132 TCAGCTCGTCAAGTGTACGCGGGGACGTACAGGATCCACTGTCCGCCAGCCCGCGGAA 191  
Db 1251 TCATCCGCGACGTGAACGTGCGGAACGTACAGGATCCAGGCCCGCGCTCCGCGAA 1192  
QY 192 CTCTGCTCTTCGCCATGATCTGTCGCGGTGTCGAGGCGAGAGGAGCGCGTAGTC 251  
Db 1191 CGCTGCTCTTCGCCATGATCTGTCGCGGTGTCGAGGCGAGAGGAGCGCGTAGTC 1132  
QY 252 CACCGCGTGGGGGTCAACCGTCCGGGTGCGCACCGGATGTGCTGCCGGGGTGAAG 311  
Db 1131 CGCGGGGTGCTCGGGAATCTGCTCCCGGCATGAACCGGAGTGTGCTGCCGGGGTCAAG 1072  
QY 312 CGCGCGCTGCTTGGCGCGGTGCTGTCGACACCGCAGGAGACCGAGTCCGCGGATGCC 371

Db 1071 GCGGCCCTGTTGGCGGGCGTCTGTCGTACACCGCGCACACAGGTCCGCCGATGCC 1012  
QY 372 GCAGAGTTTCGTCACGGTTCGGCTCTTCCCGTCCGCGGTACGCCACACCGCTTCC 431  
Db 1011 GCAGAGTTTCGTCACGGTTCGGCTCTTGGCGGTCCCGGTAGGCCACACCGCTTCC 952  
QY 432 CTCGGCTTCAGCGAGTTTCAGAGCGCGACGAGTCCGTCGGGATGCCCTCGACGTCGC 491  
Db 951 CTCGGCGGTTCACCGGTCGAGCAACCGCGCGAGTCCGTCGGGATGCCCGCGACCTGGC 892  
QY 492 GCGCAAC 498  
Db 891 AGCGAAC 885

RESULT 8  
AOPCZA361/c 37941 bp DNA linear BCT 06-FEB-2003  
LOCUS Amycolatopsis orientalis cosmid PCZA361.  
DEFINITION AJ223998  
ACCESSION AJ223998.1 GI:2894153  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
CDS  
CDS  
CDS

1. .37941  
Location/Qualifiers  
/organism="Amycolatopsis orientalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:31958"  
/clone="PCZA361"  
<1. .1014  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.27"  
/protein\_id="CAA11791.1"  
/db\_xref="GI:2894184"  
/db\_xref="GOA:O52816"  
/db\_xref="SPTREMBL:O52816"  
/translations="DHEFTTRPQFTHAKSDAHEVAQFVQGI:STYDPPHTRLRKMLT  
PEFTVRIRMEPAIOGLIDRLDMVEAGPPADLQGLFADPVGALACELGIPRDD  
QREFFVRIRENTOLSLGKLRAADSAAFNRYLDNLISROKDPDEGLGMIYREHGDN  
VTBEEKGLTALILGQVETVAGMIGCVLALLNPDQIQLLAFAPGPKADRVNVLVR  
YLSFVQAPNRLAKVVIDGQILKAGDYLCVLMANRDEALTNPVNLNANRAVS  
DVGFHGHIYCVGAALARSMLRWAYOTLMORFGLRLAVPIAEVKYRSARFVDCPDQVP  
VTW"  
1076. .2551  
/note="similar to non-heme oxygenase/halogenase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.26"  
/protein\_id="CAA11780.1"  
/db\_xref="GI:2894173"  
/db\_xref="GOA:P96557"  
/db\_xref="SPTREMBL:P96557"  
/translations="MSVDEDFDVVVGAGGPGAGSTVATLVAMQGHVRLLEKPEPPYQOI  
GESLLPATVHCVCHEMLGITDELANAGPVKRGSGTFRKGAPEPEWTTHFGISAKWAGST  
SHAYQVTRARFDELLNNAKKGVVYREGSPVTDVEDGERVTGLAYTADAGNEREVS  
ARFVIDASGNKRLYSKVGSGRNYSEFFRSALFGLYFEGGKRLPAPVSGNLSVAFDS

GMFWYIPLSDTLTSGAVVRBDAEKIOGDKALNALIAECPLISXYLANA:TRVTTG  
KYGLRVKDYQSYQQTYYRPGMILIGDAACFDVDFPSSGSHLATYLSALLAASINSV  
LAGDLDEKALNEFEMRYRREYGVFYEFVLFYFOMNVNNEESYFWQAKTKYTVQKQADIE  
SFVELTIGVSSSGTALTAAARSFAAAVDQASGDDNMFNFKSTVVRQAKAQ  
EACQVQMKALLIGEDAEPELPLPGGLVTSPDGMKMLPHHPA"  
2596. .3786  
/note="similar to glycosyltransferase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.19"  
/protein\_id="CAA11774.1"  
/db\_xref="GI:2894167"  
/db\_xref="GOA:P96558"  
/db\_xref="SPTREMBL:P96558"  
/translations="MRVLIITGCSRGDTPLVALAARLRELGAARMCLPPDYVERCA  
EVGVMPVYGRVAVRAGAREGELPGCAAEVUTEVVAEWPDKVPAALIEGCDVAVTTGLI  
PAVAVRSMAEKLGIPYRTVYLPDHLPEQSQAERDMXNQADRLLFGDAVNSHRASI  
GLPVEHLYDYGTDQPLWLAADPVLSPRLRDTGTVQTKANWILPDERPLSAELEFLA  
AGSTPVYVGGSSSRPATADAAMAKAVASGRRIVLRSRWADLVLPDDGACDFVVG  
ENVLQELFGRVAAAIHDSAGTTLAMRAGIQIVVVRVVDNVTEQYAHADRVARLGV  
GVAVDGVPVITDLSLSAALDTALAPEIRARATTVADTIRADGTTVAQALLFDVAVLEKP  
TVPA"  
3821. .5044  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.20"  
/protein\_id="CAA11775.1"  
/db\_xref="GI:2894168"  
/db\_xref="GOA:P96559"  
/db\_xref="SPTREMBL:P96559"  
/translations="MRVLIATCSRGDTPLVALAVRDLGADVRMCAPDCAERLA  
EVGVPHVPGSPARAPIQRAKPIAEVVRPTTEATATQDEIPAAEGCAAVVTTGL  
LAAAGVRSVAEKLGIDYFAFHCPSYSPFPPLPGPSPQDTIDIPAQERNNQ  
SAYQRYGLNSHRDAIGLPPVEDIFTGYTDHPWVAADPVLAPLQPTDLDVQGTGW  
ILPDERPLSAELEFLAAGSPPVYLGFSAGPGCIDDAARVAIBAHGRRIVLLSG  
DLVLPDDGACDFVDEVLQVLFPSAAAIHSGSAGTEHLATLGIPTQIVIPRHTD  
QPYAERVAEDIGIGVLEGGPVFTDMSAAVATALAPETRARATAVAGTITDGAANA  
ARLLDVAVSREKSAVLA"  
6495. .7721  
/note="similar to hydroxylase"  
/codon\_start=1  
/transl\_table=1  
/product="PCZA361.22"  
/protein\_id="CAA11777.1"  
/db\_xref="GI:2894170"  
/db\_xref="GOA:P96561"  
/db\_xref="SPTREMBL:P96561"  
/translations="MSVTSCQRI CDGTVKEIFDGRQPLSDAFVVPVGGKGFPPFLA  
TGI CDSCCTMQLMEVPEPDLMEHDAYPVLSSGSAFMRTHFNELAKRLITLTGTGDPF  
IVELGCDNGIMLKMAEAGVQLGVFSGSVADLAAGKIVRKDPFEETATADIRET  
DGPADVIYAANTLCHI PMDSILKGVTTLLGPNVGTVPFEDPILGDIVERTSPDIYDE  
HFFFTASVQMEKMAKNGLELVVERIPVHGKRVYVYGATAKATATVNFCTGTP  
DLVEFISDTTAKQNRILSPGQHI PVREPKEFAADYPDYVALLFAWNHADIINNAQAFR  
DAGGQWILYVNVNHS"  
7732. .8556

Query Match	43.2%	Score 216;	DB 1;	Length 37941;
Best Local Similarity	75.0%	Pred. No. 1.4e-19;		
Matches 270;	Conservative	0;	Mismatches 90;	Indels 0;
Gaps	0;			
QY	140	TCAGGTGACGCGGGGACGATACAGGATCCACTGTCGCGCAGCCCGCGGAGAACCTTCCTGCT	199	
DB	7716	TCAGGTGACGCGGGGACGATACAGGATCCACTGTCGCGCAGCCCGCGGAGAACCTTCCTGCT	7657	
QY	200	CTTTCCGCGCATGATCTCGTCGCGGTTCAGGCGGAGAGAGAGCGGCTAGTCCACCGCGT	259	
DB	7656	CGGCGTTCATGATCTCGTCGCGGTTCAGGCGGAGAGAGAGCGGCTAGTCCACCGCGT	7597	
QY	260	CGGCGGTGAAACGCTCCGGGTCGCGCAGCGGAGTGGTCCGCGGGGTGAGCCGGCCCT	319	
DB	7596	CGGCGGTGAAACGCTCCGGGTCGCGCAGCGGAGTGGTCCGCGGGGTGAGCCGGCCCT	7537	
QY	320	GTTTGGCGCGGCTCGTTCGACACACCCAGGAGACAGGTCCGAGCCGAGTCCGCGAGT	379	
DB	7536	GTTTGGCGCGGCTCGTTCGACACACCCAGGAGACAGGTCCGAGCCGAGTCCGCGAGT	7477	
QY	380	TGCTCACGCTGGCGCTTCCTCCGCTCGCGCGGTAGCCACACCGCTTCGCTCGGCT	439	
DB	7476	TGCTCACGCTGGCGCTTCCTCCGCTCGCGCGGTAGCCACACCGCTTCGCTCGGCT	7417	
QY	440	TCAGCGAGTTCAGCAGGCGGAGCGGTGCGTGGGATCCCTCGACGTCGCGCGGAGAC	499	
DB	7416	GCAGCTTGGTCAGCAGCGGATCAGTCTGCTTGTATCTTTCACGTTGCTGGCAAGC	7357	
RESULT 9				
LOCUS	AME16952/c			
DEFINITION	AMycolatopsis mediterranei biosynthetic gene cluster of balhimycin.			
ACCESSION	Y16952			
VERSION	Y16952.2			
KEYWORDS	btgA gene; btgB gene; btgC gene; bhaA gene; bhp gene; bpsA gene; bpsB gene; bpsC gene; bpsD gene; dihydroxyphenylacetic acid synthase; dpgA gene; dpgB gene; dpgC gene; dpgD gene; enoyl-CoA hydratase; enoyl-CoA-isomerase; glycosyl transferase; halogenase; hydrolase; hydroxylacyl-dehydrogenase; ORF1; ORF10; ORF11; ORF2; ORF3; ORF5; ORF6; ORF7; ORF8; ORF9; orfX; oxaA gene; oxaB gene; oxaC gene; oxaD gene; P450 monooxygenase; peptide synthetase; pgat gene; phenylglycine amino transferase.			
SOURCE	AMycolatopsis mediterranei			
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Mycolatopsis			
REFERENCE	1			
AUTHORS	Pelzer, S., Susemuth, R., Heckmann, D., Recktenwald, J., Huber, P., Jung, G. and Wohleben, W.			
TITLE	Identification and analysis of the balhimycin biosynthetic gene cluster and its use for manipulating glycopeptide biosynthesis in Mycolatopsis mediterranei			
JOURNAL	Antimicrob. Agents Chemother. 43 (7), 1565-1573 (1999)			
MEDLINE	93318579			
PUBMED	10390204			
REFERENCE	2			
AUTHORS	Pfeifer, V., Nicholson, G.J., Ries, J., Recktenwald, J., Schefer, A.B., Shawky, R.M., Schroder, J., Wohleben, W. and Pelzer, S.			
TITLE	A polyketide synthase in glycopeptide biosynthesis: the biosynthesis of the non-proteinogenic amino acid (S)-3,5-dihydroxyphenylglycine			
JOURNAL	J. Biol. Chem. 276 (42), 38370-38377 (2001)			
MEDLINE	21486455			
PUBMED	11495926			
REFERENCE	3			
AUTHORS	Recktenwald, J., Shawky, R.M., Puk, O., Pfennig, F., Keller, U., Wohleben, W. and Pelzer, S.			
TITLE	The nonribosomal biosynthesis of vancomycin-type antibiotics: A heptapeptide backbone and eight peptide synthetase modules			
JOURNAL	Unpublished			
REFERENCE	4			
AUTHORS	Pelzer, S.			

/codon_start=1	
/transl_table=11	
/product="PCZA361.23"	
/protein_id="CAA11778.1"	
/db_xref="GI:2894171"	
/db_xref="SPTREMBL:O52804"	
/translation="MSQNLGAGLLAISPHLDVAIVSFGAGLARAQDGAKVTVTVTF AGTATPSPAAERHGIWGLSPDASLHRRNEDIAALDHLGVYRHRFLDAIYRT LPRGLADVPGKQALSLSPQDTPDLFAVRDDIKSVVEEDQPTLLITCAAGNG HINETRDALLVAHEKLPVLMEDLPHAMFGAPAEPLPSGDLGTGDFGSGVITDM RDRKFAALRIYPSQMLHGFQKDFQAQDEHAKNSPQGGIGETWFPVVSREDS"	
8589..9431	
/note="similar to methyl transferase"	
/codon_start=1	
/transl_table=11	
/product="PCZA361.24"	
/protein_id="CAA11779.1"	
/db_xref="GI:2894172"	
/db_xref="GOA:O52805"	
/db_xref="SPTREMBL:O52805"	
/translation="MSQNLGAGLLAISPHLDVAIVSFGAGLARAQDGAKVTVTVTF AGTATPSPAAERHGIWGLSPDASLHRRNEDIAALDHLGVYRHRFLDAIYRT LPRGLADVPGKQALSLSPQDTPDLFAVRDDIKSVVEEDQPTLLITCAAGNG HINETRDALLVAHEKLPVLMEDLPHAMFGAPAEPLPSGDLGTGDFGSGVITDM RDRKFAALRIYPSQMLHGFQKDFQAQDEHAKNSPQGGIGETWFPVVSREDS"	
8589..9431	
/note="similar to aminotransferase"	
/codon_start=1	
/transl_table=11	
/product="PCZA361.25"	
/protein_id="CAA11780.1"	
/db_xref="GI:2894183"	
/db_xref="GOA:O52815"	
/db_xref="SPTREMBL:O52815"	
/translation="MSQNLGAGLLAISPHLDVAIVSFGAGLARAQDGAKVTVTVTF AGTATPSPAAERHGIWGLSPDASLHRRNEDIAALDHLGVYRHRFLDAIYRT LPRGLADVPGKQALSLSPQDTPDLFAVRDDIKSVVEEDQPTLLITCAAGNG HINETRDALLVAHEKLPVLMEDLPHAMFGAPAEPLPSGDLGTGDFGSGVITDM RDRKFAALRIYPSQMLHGFQKDFQAQDEHAKNSPQGGIGETWFPVVSREDS"	
8589..9431	
/note="similar to haloperoxidase"	
/codon_start=1	
/transl_table=11	
/product="PCZA361.30"	
/protein_id="CAA11784.1"	
/db_xref="GI:2894177"	
/db_xref="GOA:O52809"	
/db_xref="SPTREMBL:O52809"	
/translation="MSQNLGAGLLAISPHLDVAIVSFGAGLARAQDGAKVTVTVTF AGTATPSPAAERHGIWGLSPDASLHRRNEDIAALDHLGVYRHRFLDAIYRT LPRGLADVPGKQALSLSPQDTPDLFAVRDDIKSVVEEDQPTLLITCAAGNG HINETRDALLVAHEKLPVLMEDLPHAMFGAPAEPLPSGDLGTGDFGSGVITDM RDRKFAALRIYPSQMLHGFQKDFQAQDEHAKNSPQGGIGETWFPVVSREDS"	
8589..9431	
/note="similar to peptide synthetase"	
/codon_start=1	
/transl_table=11	
/product="PCZA361.18"	
/protein_id="CAA11773.1"	
/db_xref="GI:2894166"	
/db_xref="GOA:O52803"	
/db_xref="SPTREMBL:O52803"	
/translation="MSQNLGAGLLAISPHLDVAIVSFGAGLARAQDGAKVTVTVTF AGTATPSPAAERHGIWGLSPDASLHRRNEDIAALDHLGVYRHRFLDAIYRT LPRGLADVPGKQALSLSPQDTPDLFAVRDDIKSVVEEDQPTLLITCAAGNG HINETRDALLVAHEKLPVLMEDLPHAMFGAPAEPLPSGDLGTGDFGSGVITDM RDRKFAALRIYPSQMLHGFQKDFQAQDEHAKNSPQGGIGETWFPVVSREDS"	
8589..9431	

Direct Submission  
 Submitted (24-MAR-1998) S. Pelzer, Universitaet Tuebingen,  
 Lehrstuhl Mikrobiologie-Biotechnologie, Auf der Morgenstelle 28, D-72076 Tuebingen, FRG  
 revised by [6]  
 5 (bases 1 to 56917)  
 Pelzer,S.  
 Direct Submission  
 Submitted (03-AUG-2001) S. Pelzer, Universitaet Tuebingen,  
 Lehrstuhl Mikrobiologie-Biotechnologie, Auf der Morgenstelle 28, D-72076 Tuebingen, FRG  
 On Aug 9, 2001 this sequence version replaced gi:3688113.  
 FEATURES  
 source  
 1..56917  
 /organism="Amycolatopsis mediterranei"  
 /mol\_type="genomic DNA"  
 /strain="DSM5908"  
 /db\_xref="taxon:33910"  
 1..9498  
 /gene="bpsA"  
 /gene="bpsA"  
 1..9498  
 /gene="bpsA"

gene  
 CDS

/function="involved in the biosynthesis of the balhimycin  
 heptapeptide backbone"  
 /citation=[3]  
 /codon\_start=1  
 /transl\_table=11  
 /product="peptide synthetase"  
 /protein\_id="CAC48360.1"  
 /db\_xref="GI:15131492"  
 /db\_xref="GOA:Q93921"  
 /db\_xref="SPTREMBL:Q93921"  
 /translation="MNSAARTPTWLDLPSHVDRTPDAVAVAGDGVLTYYRQDRA  
 GRLAGLASGTERGRDVAIVMDRSADLVALLAVKAGRAIYVPVDAGYAPRAVAFV  
 AUSAALVVCSSASGAVPAGWMEFGEAALHSPHAFDASLFELMDPLAGVAVLLA  
 EPGSDVRRLEAAAGVTRVLTAGSLRAVABEAPESFAFREVLTGCDVPAHVAE  
 RVRTAAPARFRNMGPTTEATWHLQPDGVVGPVPTGRPLTGRVQVLDASLR  
 PVGPDGVGLVADGAEYFNRAALTAERFVADPSAQRMVGTDLQWATADGELV  
 FAGRADVDKIGFIERGEEIEAALIAPQDVHDAVAAVDGRLIGVYTESGDARVLI  
 RELIAGVLEHVAIVLALDALPTGKGVDRSALPAPEASNAGSAPSTDAEVL  
 CGLFAVLGVARAGVDGDFELGSDSIGMRLAARAAKALLVTPAQIPEPTPARLA  
 AVARFPAGPVDGELLTAAAEALAPGABEIMPLAQLQGLFESLIDDQGS  
 LTVQVQLELNGDVAAPRAANDAVRRHPELPSFLHSLAGTKVQAVAGDVTPW  
 RYDVLTAGDVDAVAALVAEQQRFEATAPLVRLVLRADRYELLVTHIHL  
 DWSAVILNDSYEAEGEPVPEQQRGATFRDYLAWLDRODDAAWAAWELAGLDR  
 BGLAVLTSVETDYRATHLPALHTLLGLFARHGLTPSTVHAANWALRLTR  
 TVYGTGWTATPEBLGIESMPGLMTAVPVRLDGGQVLDMLTLHSRQALXR  
 HOYLGLPELQKAGCATDMLVVENTPREYARTHLRTIEGTHYPTVLTGTPGDR  
 FKIQGYWPGQVPTVAESLLEWFGAICAGVAGLVLGRIGMGAADVRRWDPLQA  
 GEPLVGRMAARPDPNVAVVDGGLSADLWERSLKFAAVLAHGVRSDEDRVGLV  
 VGRSAWTVGLMGLVLLAGGTFVPVDPAPRAKEMI FRSAFVVCAGATRGVPAE  
 PADRLVVIDEVPDPAAGSGLDPRVPSAAVITYSGSTGTPKGVVTHAGLGNLAL  
 HIDRFVSPSSRVLOFALGFDTIYSEVMALLSGATLVPPRDLPRASFTDALER  
 WDTTHVKAPSVLTADVLPTSTETVVAAGELCPGLVRLSADREMNINAGPTETI  
 CATMNSPLPGQHPIDPFKPVFVGKGLLDSEFLRPLPGVGTGELYIAGIGVARGTLGR  
 SALTARFVDFPVGEMRYTGDLYAYTEQGLSVAGRADDQKIRGFVRPEIEIF  
 AUSGPRVTVQAAVADDRDLVAYTPGDIDTQAVRAHLASRMPQVMAVAVDALP  
 LTAHGKIDARALPDPDTAGKQAREPATETERVLCFLGLGLARVGVDDSDFFELGG  
 DILSNOLARAGRSGLTPTAADVDGKTPERIAQIAAESSVPEKPSKPDGVDDVA  
 WTPVMMLGDGAGAFACOMVVTGPTSDLTEKALAGPAAVVDTHMLRARVAVDEGG  
 RELVGERGSDVAGAVTEIRADGSLDEAVADAARAATRLDPSAGYACAVMDAG  
 PQVGRVLVVAHLSVDGVSRWILLSLQACAEAAVAGREPVLEFVPSGAFKRWAGLLA  
 ENVASRAGEALAAWAKLILGPDGPAGATQASRAAGVRSRVSVPKVEATLAGRA  
 PVAFHGVNVELLAGAVARWGGDAVLVDVSGHRHPVDGTLSTVGNFTSAHP  
 VRLDVADLADVLGAGPAAGRLKAVKEQSAVPDGDGLGGLRLYLNCTGTVNSHDL  
 PQPQIENQMGFRFAEKGRVAVQVGDIGSSLEPSGMLPHALBNVAI VODLPDGE  
 LTLMLFQWGLGDEKIDRLGRAWLDMLSGVARQADPAAGCTGSDPDLTLQDAE  
 EALAEFAAGAGLAVLPLSLQHGGLAFHAGVAGVDVYTTAQAVLELAGLDVPLLR  
 KSVRALDRHMLRAGFRGADGTAQVVPVAVVPTLVLDVIESADFAAGAAVAAA  
 ERARPELDRAPPLLRVMMVVLGDRHRLVLTNNHILLDGLWSTPLLDLLELLTYRNGAA  
 PALAAETPYRDLVRETREATAEAWDRALAGLPEPTLVAAADRPVPEVEPQIWT  
 TLDETAPQALGARARECGVTSTVTLQAVWGMVLAALTGRDDVDVFGSVVSGRPAELP  
 GI

gene  
 CDS

gene  
 CDS

ETMVGLEINTVPVRVVRMRFQDTPAEFLVRGLONEQVALLAHVHVLTDIOQAAGLRLFL  
 DTIVVENTYPRAEIGDESADARVRVQGLTADATHYPLAALAVFETDRLRLRLEHQP  
 ALTPAQAGAVLERFTLVLEAVVADPRLPLAVVPLIADSHYRQLOAGNTOALRLEHQP  
 LPLFAQAATAETAVVFEDESLVYAELEDAANOLARWLIDQAGAEGLVALLPLR  
 SLELVALLAVNTKGTGANPLIDPGYPADRIAKMLDDAGNALVITTAFLVASPIGDVLA  
 ARSKTVLDEPAAAGQAGDRAPVTDTRARALDDRHEFAYLIYTSGSTGRPRKAVVT  
 HRNLTNLLHCGRMYPLGRSVLHSSIAFDLTVTATFTPLIVGGSHVGLADLILGV  
 VEAPSIPLIKATPESLITLDTASRAGSGLDILLGEGQLPADTVQWRKYPNIVVN  
 EYQGTATVGVETRLQEQCPGPGVPTIGTPLANRAFLVLDWSLRLVPPGAVGLY  
 GAGLARGVLRAGLTGTFRVADPFGSGRMVYTGDLVQWNPDGQLVFPAGRVDDVVK  
 VRGFIEPGRIBALVAQSVGOAVVARDSEIGTGLIGVYTAAGSGVDVEAAVREGV  
 AERLPOYMPAALVVLGALPLTANGKVDRAALPDPPDGRAGRGREPTAEERLICALF  
 AEVLGERAGADDSFELGDSILSNRLAARAREGSGFARVFPQRTPAGTAAIVE  
 RVAGDRPAAVHAVSVDALLDQGLDEFKAEFDSDSQFFADPGRY"  
 9498..21731  
 /gene="bpsB"  
 9498..21731  
 /gene="bpsB"  
 /function="involved in the biosynthesis of the balhimycin  
 heptapeptide backbone"  
 /citation=[3]  
 /codon\_start=1  
 /transl\_table=11  
 /product="peptide synthetase"  
 /protein\_id="CAC48361.1"  
 /db\_xref="GI:15131493"  
 /db\_xref="GOA:Q93920"  
 /db\_xref="SPTREMBL:Q93920"  
 /translation="MNSQRIEIEWPLSQAGLLFHAVVDGEGDVPYVIGHWILDLAGP  
 VDAAGLRAAETELLARHAPLACFRQKSGETVQIIARQVELPWRVDSHLDLDPSEA  
 VRELAEQDRTTFDLAQALLRLIRLGADARHVTVCHTHTMDGWSLPIVLDLSL  
 LYPAGDASALPDVPSRYEAWLSQDKERALSAAWLSGABEPTLVVPAVGRAPR  
 AYESVEAHLPEHLTRSLAEAPRHGLTINTVQGAALVLAOLAGRPDVFQGAASV  
 RPDLPQTEGMGLFINTVPVRLSGSPVVELLAELKROQSALIPOCFVGLADIQQ  
 AAGPAVFTLLVFEKHFHGPAGSDAGTFRIHVNOGRVAAHTPLILVAPGSMILK  
 LDYLTEFURETAPAILERTVLRQLTGAGELTVAGVEVTTAAERLAVAGEGASTS  
 APPSLPALDFGHQVARRDEPAVDGDTVSYGELAEKRLADSGPAAVCAEYRAAV  
 AVTLDSPLDIATLLAVKAGAVVPVDPAYPVERKFKMLADSGPAAVCAEYRAAV  
 PDCPEPVLDDPRTQAVAESPRISAGTSADLLAVMTVTSSTGTGPKGVAVSHGNA  
 ALAGEPWRVPGCDVLAHSHAFDILSEFMVPLLSGARVVLAVGPAVGAALAAV  
 AGVTAHHTAGAPVRLADESPAVAGLEVLGSDVAPLAAVERVGRVGRVVRHL  
 YGTEATLCTATWLLPEGBETGSLVPIGRLAGRVHLDALFLRPVPPGVAGELYVAG  
 AGVAGQSPALTAERFVADPSGSGARMTGDLATWTEQGALLAFGRADDQVKIRG  
 YRVEPEIEVITVLAGLVQAVVTTPRGEHLIGYVVAEGHADDPVRLBQLAGTLPEF  
 MPAAVLLDELPLTVNGKVDRLALPEPFAKSGREPVEARVLCGFVADVLGID  
 HVGVDSDFFELGDSISMSQVAAARREGISLTPRLVFEHRTPERLAAALQAEGATPR  
 AEVVTGPELTPMTVMRALGDDAMRPGFAQVVRVVTAGVPDVALYALQADLADHL  
 LRAVPPDGLVLPESGAVAAACLLTRVAGTGLDEIAEREVTRTATGLDPSGJMA  
 RVVMDAGDEPGRLLAFVAHLSVDVAVSWGILLPDLRAAYDEVISGTPALEPPTVS  
 RQWARRLTALSESITVALEKAAVVEGEPALPQDTQHTQCSHSWSTSLSTETVR  
 DLVTLPFGACGIDVLLAGLAGAVRVGSGAALLVDVEGHRSAADGEDLLRTVG  
 WFTSVHPVRLSDVLDAGADGERPAGQLLKAVKEQIRAVPDGSGYGLLRHNPOT  
 GARLELPSAQIGFNYLGRVLAPEPTAMQNGGPGPLGGPDMVLAAHVEVSAEQDT  
 PAGRLGLAIDTRDFLATVERGEAWLEMLTGLAAVARGSGAGGHTPADFALVDLTQ  
 RDVALEAAAGLTDIWLPSLOSGMLFERADGVDVYVQTORIILDQGLDEPLRL  
 ANQVLRHASLRTGFHQLGSGATVQVVRADIPRVADILSHLDAEAAAEPLRL  
 AEOGGRFVTRPRLRLRLIRLGDEHRLVTVSHVLLDGMSTPLVUGMSDGYAGG  
 RSSKSPSYQDYLASLRQDAETASRAARLAGEDEPTLVADAGATLVMPDEHAEV  
 LPEFATRALGAFARGHLTVSTVLGAWALVTLARAGTDDVFGSVGSPADVPDVE  
 RMVGMINTVPARVLDGRRLLEMLDQARQALTEHQYLGLEPTIKVAGTCAIFD  
 TIVMENVYHDAAGLGGDGVATSSVTRTGTSTPLTMVSLGDLRLITVSYRPRDI  
 DAAEAVAROVRLVERVAPSLVGRIGTSETPTAAVVERWNSTGEAAATSVLE  
 LFRQAGASDAVAVVAGERTLSYADLDRSDRLAGLAGIGVGRDGVVMTGRAD  
 LFVALLGVWAGAAQVFNVDYPAERIEMLAUVGASVAVCEATRAKVPDGPVPPV  
 DLPVIGVGRPEAFPTVGAHDVAVVMTVSGTGVKAVAVPHGVSAAALADPGWQGP  
 GDCVLLHSHAFDASLVEIWPVLSGARVLVAEPGTVDVREARVSGRTVTHLZAG  
 AFRAEASPDPSLIGRLITGGDAVPLASVVRMQACPDVVRQYLGPTETILCATW  
 LVLEPGAAGDVLPIGRPLAGQAVYLDLQVPAVNVNTELVAGLHAGLYLGNTA  
 ATSERFVAVPESGGRMYRTGDLARWTDQGLVFRAGADSQVKIRGVVRPEPVEVAV  
 TEPHVAQVAVARECOPGEKRLIAYTAAEAGSALSAAYRAHLATRELPEVPSVV  
 VLDSFPULLNGKIDRALPAPEFAGKAGREPTBAERVLCGLFALILGLERVGADG  
 FPELGDSDILSLMELARENFVFGAKQVFEQKTPAGIAAARVAGRGQSAPGAVGSR  
 GEVFTVPVTRALLERDPAGLTRGAAQWVSVAAPRDLSTVTLVAGLGAVIDTHDMLRS



REFERENCE AUTHORS	TITLE
JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	TITLE

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	REMARK	JOURNAL MEDLINE PUBMED REFERENCE AUTHORS	COMMENT FEATURES	sound	gene
--	-------	--	-------	--	--------	--	---------------------	-------	------

gen  
CDS



```
/translation="MTTTSILVTGGAGFIGSHVYRTLLGPRGVDDVTVTVDLKLTYAG
TITNLAETSDSRFRFFVRGDI CDAPLDVDDLLAVHQVVFHFAAESHVDRSILGAADFVR
TNVGTQTLLDAALRQGIETFVHI DTDEYIGSIDAGSWPETAAPSNSLYSAAKASD
LVALVYHRTGHDVTRCSNNYSGHOPPEKVIPLFVTSLLDGREVLYGDTGVURDW
LVVDHVRALVTRVTCGRAGEVYNTGGTELSNKELTQLLLDAGAGHVRVYVTRDK
GHDRIYSDCTKIRBELGYRPAREFGDALAEIVANRYRHRHAWPEPLIRYCAVAA"
complement(2604..3365)
/gene="mtmrii"
/complement(2604..3365)
/gene="mtmrii"
/codon_start=1
/transl_table=11
/product="ketoreductase"
/protein_id="CAA07756.1"
/db_xref="GI:3334821"
/db_xref="GOA:O86480"
/db_xref="SPTREMBL:O86480"
/translation="MDMGLTGRRLVLTGGSGIGAAVAVADEGARVALTVRGDEA
ARLARELAGRHAAMAVRYSIGEPTVESAVATVTERWGGVDVIVANAVWAPRRAP
GTRFEVASGDRNGRFLDNLAPTLRIVQLVYAGMEARSWGRIVLVLSHVLDLHGRQOE
FYGAASPEDIFARSLAMDVGRGVLVNVVCPGLTITRVLVTLGLDEIRERELGSTPT
GRLSPEDTAGAVFLGSAANANTTGGTLTVSGGR"
complement(3382..4977)
/gene="mtm011"
/complement(3382..4977)
/gene="mtm011"
/codon_start=1
/transl_table=11
/product="oxygenase"
/protein_id="CAA07757.1"
/db_xref="GI:3334822"
/db_xref="GOA:O86481"
/db_xref="SPTREMBL:O86481"
/translation="WADPTPLVAGAPVGLTAHRLARGRLVPLIDAAPRAATSR
AVNPHRTLETVDQMGVIEELIENLAQQIPAFNTRSGRRLVRLDADYITMTRPFFTA
IVQVDTEAVLSAVARLGDVVENETRLTGFSDAGVDVTLLEHADGITESTRVVPLFV
GCGHSTVTKLLGPLLGESSETWMLADAPVDVLPDPSIYVWHTGQALMVPPFR
GGRWLLDTTAEPTGDAAVIADRFTRNSAPASAVRQVPTPTSVFTFOQRMVPRMG
EGRVFVAGDAKATVPSKGRMNTGVQEAYNLAWLALVAEGHARELLDYSISLRVP
IGERLIGSKTATFLVKNAPAPIALPVFGLIRNVPPLREKVKORVLGIGISGLQLD
YDPSPLTRPAATGAGPRCPTRASVSGSDPADPQGCRAABELRDTFWLANTVPDVAAG
APANSWLVRILGDSQSPGLPADPNGLRLAUGLASGGVLLVRPDGTVAARGRALG
GADRLALAAAGLRDQHVLPHS"
complement(4970..5269)
/gene="mtm011"
/complement(4970..5269)
/gene="mtm011"
/codon_start=1
/transl_table=11
/product="oxygenase"
/protein_id="CAA07758.1"
/db_xref="GI:3334823"
/db_xref="SPTREMBL:O86482"
/translation="WVTFVNKLTVNGDIEEFLTIKESLTERMSAOPGYRSSTLLRQVG
RPDPVTELEMDDAASHQAARSEEFQARVKGILGALASVDPLGYIEASEDGNRG"
complement(5313..6857)
/gene="mtmL"
/complement(5313..6857)
/gene="mtmL"
/codon_start=1
/transl_table=11
/product="acyl CoA ligase"
/protein_id="CAA07759.1"
/db_xref="GI:3334824"
/db_xref="GOA:O86483"
/db_xref="SPTREMBL:O86483"
/translation="MSELPGPAPVLDLGLRLAABERHPEQAVRDERGADVYAGLDAR
ADRFARALLRTGRTAVGASVLDPVFAAFYRSRSGNRVLVNLVREPVLHVH
FRTAGIEIALVSSETSPPGBRALPLREVVVDADRTGPPPTREPLDLLLTEDGT
AALDPFAGVLDLSVVCVQTSCTGTPGKGVRLIHRNLVANAQAHAHLGDRAGVCLN
HLPLVHMLDLSAVYAGATQVLCHDPPDVASAAABAGATHYGLPLVLAARLADPR
LASVWPGELGRLSGSRLAPAAARLERLGVPIQYGLAELSPISHNDRPERSK
PGSVGPAVPGTECRIVDLETGAALDPGRPEVLLRGPQLMAGYLGLPDAPDIAAGWF
```

```
HTDGVGYQDEBGMFLVLDRIKQVFKVDNBLVSPSEIEQVLLQDPDADVCVADLPDEF
SGAVWAGVVVPGDGPVDLNPVARANALLSDHQIRRAERLTAVPSPNGKTERTL
RELRLA"
complement(6850..7797)
/gene="mtmQ"
/complement(6850..7797)
/gene="mtmQ"
/codon_start=1
/transl_table=11
/product="aromatase"
/protein_id="CAA07760.1"
/db_xref="GI:3334825"
/db_xref="GOA:Q53669"
/db_xref="SPTREMBL:Q53669"
/translation="MNSQOYSVRDEILVAAPARVRSVLVDVVGWVTLHGPAVHAHV
LRDTGDERLIRHWATVGRASVTRTARHTRSDASRISFAHVMPPEAPFASLEGAWTFR
ERDCKTRVENRHDFTLSDGSPADBLTARLREGSRAYLSTLGHAAENHEELRLIIS
FTDSTFAGVDKDAYDNLWEADKWEERIPIHIAKLDLEBPAGVQVDFMLTSTPDGSSH
RTSSVTVTPEDLIIVKQTELPALLDGHGTWKFRTTPEGVIAEASHTATVKPSALHI
LGSGTTVQDARNYLRRVLSNSVSNLRKAKAYAEERAGV"
8009..8461
/gene="mtmX"
/8009..8461
/gene="mtmX"
/codon_start=1
/transl_table=11
/product="cyclase"
/protein_id="CAA07761.1"
/db_xref="GI:3334826"
Query Match 32.7% Score 163.4; DB 1; Length 18977;
Best Local Similarity 66.8%; Pred. No. 1.1e-12;
Matches 233; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 151 CGGGGGAGGTACAGGATCCACTGTCCCGCAGCCGCGGGAACCTCTCTCTCGCCATG 210
DB 15183 CGCGGGTGGCGAGGATGAACCGTCCCGCTCCGCCAGGAACGCCCTCTCTTGGCGAGG 15242
QY 211 ATCTCGTCGCGTGGTTCAGAGCGCAAGCAGCGCGTAGTCCACCGCGTCGGCGGTCAAC 270
DB 15243 ATCTCTCGCGGTAGTTCAGGCGCAGCAGGAGTAGTAGTGGCTGGTCTTGGCG 15302
QY 271 GCGTCGCGGFGTCGACCGGAGATGTGCTGCGCGGGGTGAGCCGCCCTCTCTTGGCGCGG 330
DB 15303 TACTCGGGGTCGCGACCGGATGTGCTGCGCGGCGAGCACCAGCGCTCTCTTGAAC 15362
QY 331 GTCGTGTCGCACACCCAGGAGACCAGGTCCGCGACCGATGCCGCAAGAGTTTCTCACCGGTG 390
DB 15363 GTGGTGTGCGACGAGATTCCAGGTCTGTCGCGGTGAGCCGCGAGAGCTTCAGCAGGGTG 15422
QY 391 GCGCTCTTTCGCGCTCGCGCCGTACGCCACCAACCGCTTGGCTTGGCTTGGAGCGAGTTC 450
DB 15423 TTGCCCTTGGCGGCGCCCGTACCCGCGATCCGTTTGGCTTGGCTTGGCGCGCAGCGAGCGG 15482
QY 451 AGCAGGCGGAGCAGGTGCGGTGCGGATGCCCTCGACGTCCGGCGGCGAAC 499
DB 15483 ACCATGAGGTGAGTTCGCGGTGATCTCCGCGACGTGCTTGGCGAAC 15531
RESULT 11
AR088311/c
LOCUS AR088311 1401 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5989869.
ACCESSION AR088311
VERSION AR088311.1 GI:10015074
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1401)
AUTHORS Filippini,S., Lomovskaya,N., Fonstein,L., Colombo,A.Luisa. and
Hutchinson,C.Richard.
TITLE Process for preparing daunorubicin and doxorubicin
```

JOURNAL	Patent: US 598969-A 1 23-NOV-1999;
FEATURES	Location/Qualifiers
SOURCE	1..1401
ORIGIN	/organism="unknown"
	/mol_type="unassigned DNA"
Query Match	30.0%; Score 149.8; DB 6; Length 1401;
Best Local Similarity	65.3%; Pred. No. 1.3e-10;
Matches 220; Conservative	0; Mismatches 117; Indels 0; Gaps 0;
QY	164 GGATCCACTGTGCGGCAGCCGGCGGAACCTCTGTCCTTCCGCCAATGATCTGTCGGCGT 223
DB	1390 GGATCCAGCGGCCCCCTCTTCGTGAATCCCGCCTCTCGGGCCTTGGAATCTTCGTCTAGGT 1331
QY	224 GTTTCCAGCGAAGACGACGCGTAGTCCCAACCGCTCGGGCGTGAAACGCTCCGGGGTGC 283
DB	1330 GTTTCCAGCGAAGACGACGCGTAGTTCGGGTAGGGGCGCCGGAATCTCTCGGGGAGC 1271
QY	284 GCACCGGGATGTGCGTGC CGGGGGTGAAGCGCGCCTCTGTCGCCCGCGTCTGTGCGACA 343
DB	1270 GGATCGGGATGTGGGACCGCGGGAGCGGACCGCGCATCTTGGCGCGCTGGTGTGCTAGA 1211
QY	344 CCACGAGACCGAGTCCGACCGATCCGACAGTTCGTACGCTGCGCTCTTCGCGCG 403
DB	1210 CGCACGGCAGCATCGGGCGCATCCCGCATAGTTGAGCACTGTGGCGCTCTTCGGCG 1151
QY	404 TCGCGCGGTACGCCACACCGCTTCGCCCTCGCGCTTGAAGCGATTGAGCGAGTTCAGCAGGGCGACA 463
DB	1150 TCGCGCGGTACGCCACACCGCTTGAAGCGAGTTCGCGCAGCTTCGCGCAGACGGGGCACCA 1091
QY	464 GTCGCTGCGGATGCCCTCGACGTCCGGCGCGAACCT 500
DB	1090 GTTCACGACGACCGCTTGACCTGGCGCGCGAACCT 1054
RESULT 12	
AR144738/c	
LOCUS	1401 bp DNA linear PAT 08-AUG-2001
DEFINITION	Sequence 4 from patent US 6210930.
ACCESSION	AR144738
VERSION	AR144738.1 GI:15106605
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unclassified.
REFERENCE	1 (bases 1 to 1401)
AUTHORS	Filippini,S., Lomovskaya,N., Fonstein,L., Colombo,A.Luisa,, Hutchinson,C.Richard., Orten,S.I. and Brame,U.
TITLE	Process for preparing doxorubicin
JOURNAL	Patent: US 6210930-A 4 03-APR-2001;
FEATURES	Location/Qualifiers
source	1..1401
	/organism="unknown"
	/mol_type="unassigned DNA"
ORIGIN	
Query Match	30.0%; Score 149.8; DB 6; Length 1401;
Best Local Similarity	65.3%; Pred. No. 1.3e-10;
Matches 220; Conservative	0; Mismatches 117; Indels 0; Gaps 0;
QY	164 GGATCCACTGTGCGGCAGCCGGCGGAACCTCTGTCCTTCCGCCAATGATCTGTCGGCGT 223
DB	1390 GGATCCAGCGGCCCCCTCTTCGTGAATCCCGCCTCTCGGGCCTTGGAATCTTCGTCTAGGT 1331
QY	224 GTTTCCAGCGAAGACGACGCGTAGTCCCAACCGCTCGGGCGTGAAACGCTCCGGGGTGC 283
DB	1330 GTTTCCAGCGAAGACGACGCGTAGTTCGGGTAGGGGCGCCGGAATCTCTCGGGGAGC 1271
QY	284 GCACCGGGATGTGCGTGC CGGGGGTGAAGCGCGCCTCTGTCGCCCGCGTCTGTGCGACA 343
DB	1270 GGATCGGGATGTGGGACCGCGGGAGCGGACCGCGCATCTTGGCGCGCTGGTGTGCTAGA 1211
JOURNAL	Patent: US 598969-A 1 23-NOV-1999;
FEATURES	Location/Qualifiers
SOURCE	1..1401
ORIGIN	/organism="unknown"
	/mol_type="unassigned DNA"
Query Match	30.0%; Score 149.8; DB 6; Length 1401;
Best Local Similarity	65.3%; Pred. No. 1.3e-10;
Matches 220; Conservative	0; Mismatches 117; Indels 0; Gaps 0;
QY	164 GGATCCACTGTGCGGCAGCCGGCGGAACCTCTGTCCTTCCGCCAATGATCTGTCGGCGT 223
DB	1390 GGATCCAGCGGCCCCCTCTTCGTGAATCCCGCCTCTCGGGCCTTGGAATCTTCGTCTAGGT 1331
QY	224 GTTTCCAGCGAAGACGACGCGTAGTCCCAACCGCTCGGGCGTGAAACGCTCCGGGGTGC 283
DB	1330 GTTTCCAGCGAAGACGACGCGTAGTTCGGGTAGGGGCGCCGGAATCTCTCGGGGAGC 1271
QY	284 GCACCGGGATGTGCGTGC CGGGGGTGAAGCGCGCCTCTGTCGCCCGCGTCTGTGCGACA 343
DB	1270 GGATCGGGATGTGGGACCGCGGGAGCGGACCGCGCATCTTGGCGCGCTGGTGTGCTAGA 1211

/translation="MEPNSTCRICGRRVREFFDGRQPLSDYFSPSEELDNEFFRL  
AVGMCVTVMQLEEVPRDRMFRYDYPVHSSSRMRHEFAATARRLIGTELGRDP  
FCVEIGNSDVMRLTRVDAGVRHLGVSPGVADVRAGELQVTRFAFFESTAREIAQ  
EHGPANVIAANTICHI PYLDSVFRGIDALLAPDGVFVEDPILGDI VEKNTFDOYD  
EHFYLFTASVSTTIAHQFGEIVLVDVERLPHVGEVRYTIAAGRQPSRVEGLAEB  
SRGLADLTLEKFGQVKBVCCDLVARIELRDLGVVGVGATAXATVNLVAGIG  
PILLPCVYDITPAKGRRLPGSHIFIRSAEERAFYDPYDVALFWHHLDEVQAREAF  
TKQGGRWIRSG"

## ORIGIN

Query Match 30.0%; Score 149.8; DB 1; Length 2221;  
Best Local Similarity 65.3%; Pred. No. 1.1e-10;  
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GATCCACATGTCGCCAGCCGCGCGAATCTCTGCTCCTTCGCCATGATCTCGTGGCGT 223  
DB 2210 GGATCAGCGGCCCCCTGTTGTAATCTCGCCCTCTCGGCGCTGAGATTCGTTAGT 2151

QY 224 GGTTCAGGCGAAGACAGCGCGTAGTCCACCGCTCGGCGGTGAACGGTCCGCGGTGC 283  
DB 2150 GGTTCAGGCGAAGACAGCGCATAGTCGGGGTAGGGGCCCCGGAACTCTCTCGGCGAGC 2091

QY 284 GCACCGGATGTCGTGCGCGGGGTGAGCCGCGCTCTGTCGCCGCGCTGCTGTCGCACA 343  
DB 2090 GGATCGGATGTGGACCCGCGGAGCGGACCGGCGGATCTTGGCCGCGGTGTGCTAGA 2031

QY 344 CCACGAGACCGAGTCCGAGCGATGCCCGCAGAGTTCGTACGGTGGCGCTTCGCGCG 403  
DB 2030 CGACCGGAGCATCGGGCCGATCCCCCGCATAGTTGAGCATGTGGCGGCTTCCTGGCGG 1971

QY 404 TCGCGCGTACGCCACACCGCTTGCCTTCGCCCTTGAGCGAGTTCAGCAGGGCGAGCA 463  
DB 1970 TCGCCCGTACCGGACGAGTAGAAGCCGAGGTTCGCGGAGTTCGCGGAGCGGCCACCA 1911

QY 464 GGTCTGTGGATGCCCTCGACGTCGCGCGGACCT 500  
DB 1910 GGTCTGTGGATGCCCTTGACCTGGCGCGGAACTT 1874

## RESULT 14

BD251846/c  
LOCUS  
DEFINITION  
Gene cluster involved in nogalamycin biosynthesis, and its use in production of antibiotics.

ACCESSION BD251846

VERSION BD251846.1

KEYWORDS JP 2002528068-A/1.

SOURCE unidentified

ORGANISM unclassified.

REFERENCE 1 (bases 1 to 16020)

Ylihanko, K., Torkkell, S., Palmu, K. and Hakala, J.

Gene cluster involved in nogalamycin biosynthesis, and its use in production of antibiotics

Patent: JP 2002528068-A 1 03-SEP-2002;

GALLILAEUS OY

OS Streptomyces nogalater ATCC 27451

PN JP 2002528068-A/1

PD 03-SEP-2002

PF 20-OCT-1999 JP 2000578345

PR 23-OCT-1998 FI 982295

PI KRISTINA YLIHONKO, SIRKE TORKKELL, KAISA PALMU, JUHA HAKALA PC

C12N15/09, C12P7/66, C12P19/56, C07H15/252, C12P7/66, C12R1/465, PC

(C12P19/56, C12R1/465, C12N15/09)

CC 'overlapping sequence in the genes snog and snogN' CC

'overlapping sequence in the genes snog and snog' CC 'unknown

region'

FH key Location/Qualifiers

FT misc\_feature 3799..3800

FT misc\_feature 5334..5356

FT misc\_feature 13201..13300.

Location/Qualifiers

1..16020

## FEATURES

source

/organism="unidentified"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32644"

## ORIGIN

Query Match 27.0%; Score 135; DB 6; Length 16020;  
Best Local Similarity 61.5%; Pred. No. 5.4e-09;  
Matches 216; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 150 GCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAACTCTTCCTTCGCGCAT 209  
DB 6323 GCGGGGATGGGACAGATGAACCGCGCGCTGTTCGCCGAGGCCCTCTCTTGTGCG 6264

QY 210 GATCTCTCGCGTGGTTCAGCGGAGAGAGCGGTAGTCCACCCGCTGCGGCGGAA 269  
DB 6263 GATCTCTCGCGTGGTTCAGCGGAGAGAGGTAGTTCGCGGATGTTCCTTGGC 6204

QY 270 CGCTCCGCGGTGCGCACCGCGGTGTGCTGCGCGGGGTGAGCGCGGCCCTCTTCGCGCGG 329  
DB 6203 GTGCTCGGAGCGTGCACCGGTATGTGGTGGCGGCGAGCACCTGCTCTTCAGCTC 6144

QY 330 CGTCTGTGCGACACCCAGGAGACAGTTCGGACCGATCGCGAGAGTTCGTACGGT 389  
DB 6143 GGTGCTGCGGTGCGATTCAGCTCTTCAGGCGAGCCGCGACACCGGTGAGGTGGT 6084

QY 390 GGCCTCTTCGCGCTCGCGCGGTACGCCACACCGCTTCGCTTCGCGCTTCGCGCGAGTT 449  
DB 6083 GTTCCCTTGGCGGAGACACCGTACCGACGATGCGCTTGGCGCGACGAGGAGCG 6024

QY 450 CAGCAGGGCGAGAGTGGTGGGATGCGCTTCGAGTTCGAGTTCGCGCGCGAACCT 500  
DB 6023 CAGCAGTTCGCGAGTTCGCTGGCGCGCCCGCTCGATCCGCTGCGCGAACCT 5973

## RESULT 15

AF187532  
LOCUS  
DEFINITION  
AF187532 16065 bp DNA linear BCT 07-JUL-2000  
Streptomyces nogalater SnO (snO) gene, partial cds; C-7  
ketoreductase (snOAF), SnO (snO), SnO (snO), putative glycosyl  
transferase (snOE), putative hydroxylase (snOA), putative  
glycosyl transferase (snOD), SnO (snO), nogalonic acid methyl  
ester cyclase (snol), putative dTDP-glucose-4,6-dehydratase  
(snogK), putative dTDP-4-dehydrohannose reductase (snogC), Snog  
(snog), SnogN (snogN), putative polyketide cyclase (snOA),  
putative amino methylase (snogA), and putative dTDP-glucose  
synthase (snogJ) genes, complete cds; and putative aminotransferase  
(snogI) gene, partial cds.

ACCESSION AF187532

VERSION AF187532.1

KEYWORDS GI:6018298

SOURCE Streptomyces nogalater

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 16065)

Torkkell, S., Ylihanko, K., Hakala, J., Skurnik, M. and Mantsala, P.

Characterization of Streptomyces nogalater genes encoding enzymes

involved in glycosylation steps in nogalamycin biosynthesis

Mol. Gen. Genet. 256 (2), 203-209 (1997)

JOURNAL 98007868

MEDLINE 9349712

PUBMED 9349712

REFERENCE 2 (bases 1 to 16065)

Torkkell, S., Kunnari, T., Palmu, K., Hakala, J., Mantsala, P. and

Ylihanko, K.

Identification of a cyclase gene dictating the C-9 stereochemistry

of anthracyclines from Streptomyces nogalater

Antimicrob. Agents Chemother. 44 (2), 396-399 (2000)

JOURNAL 20106987

MEDLINE 10639368

PUBMED 10639368

REFERENCE 3 (bases 1 to 16065)

Torkkell, S. and Ylihanko, K.

Direct Submission

TITLE Submitted (16-SEP-1999) Biochemistry, University of Turku,

Vatsoelankatu 2, Turku FIN-20014, Finland  
 Location/Qualifiers  
 1. 16065  
 /organism="Streptomyces nogalater"  
 /mol\_type="genomic DNA"  
 /strain="ATCC27451"  
 /db\_xref="ATCC:27451"  
 /note="related downstream sequence in EMBL accession number AJ224512"  
 complement (<1..571)  
 /gene="snoN"  
 complement (<1..571)  
 /gene="snoN"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoN"  
 /protein\_id="AA01805.1"  
 /db\_xref="GI:6018299"  
 /translation="MQTEPGVADLPAESDPALRELAARVERDGVVPGVLDAE  
 VAELARARLAAHEESVRNGSGAGTMDYVDAQSGDYMRLAHPRIAAAEVLA  
 GSPRLFKLEVLLENKKEKQASVTPAPHDFAFPFSTAGTALTAWALVDVPEVGGC  
 MTFVPGSHLLPDPDTPGDEPWAGAFTRPGEI"  
 complement (644..1489)  
 /gene="snoaf"  
 complement (644..1489)  
 /gene="snoaf"  
 /codon\_start=1  
 /transl\_table=11  
 /product="C-7 ketoreductase"  
 /protein\_id="AA01806.1"  
 /db\_xref="GI:6018300"  
 /translation="PMRATDSTGPRVPVMPSPAPSPPTSPGAPGSPAPLAVITGG  
 GSGGRATAPAAQAGKVLVVGSTEDALQAQAGCADMLVADVASPDGPAVNA  
 ALRFGTRIDVUNNAVAGMETLOTYRDADVARQFGUNLAPLFLVQSALGALEKSG  
 IVNWGTATGLRAAPATGALYGASKVLDLITRNAVELAPRGIRVGVGAPGVIDTG  
 IGVKGMTPEGRFLTGMGRVPGVRGVEDVAMWIVQLARFEAGTATGMVVFVVG  
 GLSLV"  
 complement (1580..2047)  
 /gene="snoO"  
 complement (1580..2047)  
 /gene="snoO"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoO"  
 /protein\_id="AA01807.1"  
 /db\_xref="GI:6018301"  
 /translation="MSVSTDOTAAPEDRAAATDPGHLVYAOVQFVAROMQLDSCG  
 AEWAAITEDGTPRSPSPAPRCHALAGARAAERLAEGLSHEHVIGMTAVRR  
 EPDGSVFRSYAQVATPRGRGEAPRLHLICVCEDVLVREGPGIKVRVRVTHDA"  
 2267..2686  
 /gene="snoL"  
 2267..2686  
 /gene="snoL"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoL"  
 /protein\_id="AA01808.1"  
 /db\_xref="GI:6018302"  
 /translation="MSTTANKERCLEWAMNRWDVGVWAHAPDVVHYDDEKPV  
 AEVVRMNSAVEAPDLRLDVRISVIGEDRVMLRITCSATHQGVFMGIAFTGRKVR  
 TVLEELRFEAGVVEHVDVFNFSPLFRDLGVVPDGL"  
 complement (2744..4042)  
 /gene="snoG"  
 complement (2744..4042)  
 /gene="snoG"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative glycosyl transferase"  
 /protein\_id="AA01809.1"  
 /db\_xref="GI:6018303"  
 /translation="MNVLTSTFAMDAHCTAVPLAWLASAGHEVVRVACQALPTIT  
 GAGLTAVPVGDRDTHGSLLRGSDILALHDEADYLEARHDAFGFEFLKGHTVMSAL

FYSIINDSMVDDIVDFARHWRPDLVWVEPFTFAGAVARASGAHAHLLSFDPDLPLS  
 TRRLFLERMARQEEHDDTLAKWDLMTLGRGHGSHSDEEIVTGWMSIDQTPAPVRLDA  
 GGTPVMRYVPSGLVFTVVDVWMLRPERPRVLTGITSRSVKSLAVVVDLFEA  
 VAGLVQVWATLDADQRELGRVDPHFRIVEHPDLDAVLTCTCSAIVHGGAGTWSIAA  
 VYGVQVSLGSMMDHFTYARLEBELGAGLRPLSGELTAGRLTLEIRVIGLVEPSPGTA  
 QALSDTIAAPSPSEVVPVLEELTGRHPCTREPPRALRA"  
 complement (4086..4913)  
 /gene="snoaw"  
 complement (4086..4913)  
 /gene="snoaw"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative hydroxylase"  
 /protein\_id="AA01810.1"  
 /db\_xref="GI:6018304"  
 /translation="MTVLTGATGNVGRHVVTGLLAARRVRALTRTPDRSGLPGGAE  
 ITGDLTRPEYERMLDGEVAYLFPVPEATAAAGARRAGVRIIVLSSDSVTDTGT  
 DTGHRVELAVEDTGLHWHVRFGFALMKVITWASIRAGGVRSAYPARVAPVH  
 EADVAATAVALLKKEHAGRAISVTPQALTRQEQVRAVGGELGSLAFVETVQGQAR  
 ADLTAQGLPAPADIADYVLAFOAGWTERPAPAPTVREVTGRTPARTLAQMAADHADRFR"  
 5053..6225  
 /gene="snoGD"  
 5053..6225  
 /gene="snoGD"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative glycosyl transferase"  
 /protein\_id="AA01811.1"  
 /db\_xref="GI:6018305"  
 /translation="MRVPSGSCRTGIMRALFITSPGLSHILPTVLAQALRALGHEVR  
 YATGDIRAVABAGLCAYDVSGVNYAKLFVDPDQDVTDMHSEGLGSGFFPAEMFARV  
 SAVADGALRTARSRPDLVWVTTPQAGAPLTAALQICVPELPLGADSERGICALI  
 REAMQVGERHGTGVTGSPVRLTTPPSVALLPEDESRSGAMWPMRYVYNGGAVLP  
 DWLPPAAGRIARTVIGSIDALSHGGIAKLAFSDVADVDAEFVITLGGDGLALGEL  
 PANVRVWRIPLGALLETCDALIHGGSGTLLTALAAGVQCVIPHGSYQDTRDVL  
 GLGIGFDABAGSLGACQCRLLDDAGLREALLRVRQEMSEMPPTAETAAKLVALAG"  
 complement (6271..7074)  
 /gene="snoK"  
 complement (6271..7074)  
 /gene="snoK"  
 /codon\_start=1  
 /transl\_table=11  
 /product="SnoK"  
 /protein\_id="AA01812.1"  
 /db\_xref="GI:6018306"  
 /translation="MPDFGQPTAENLSKEAVRFVREQGVYHIVPLVSTETVAFRAA  
 CHEVLEKREIWDGDEQVHYVQAQWQKPELSLVHPEISGLARLAGAPLRY  
 YSDILVKEPKRTPLTVHDDDETGLPLNELSATLTANIALTDVPTVERGCMQSVPGSHL  
 RAREDRQEMTSFAEFDLADVWPDYWPQPRVAVPVAGDVVFFHHCRTVHMAENISD  
 SVMAHGVMYMDADATYRPGVQDGHLSRLSPGDPLEGEFLPLVTAGTQ"  
 complement (7125..7529)  
 /gene="snoAL"  
 complement (7125..7529)  
 /gene="snoAL"  
 /codon\_start=1  
 /transl\_table=11  
 /product="nogalonic acid methyl ester cyclase"  
 /protein\_id="AA01813.1"  
 /db\_xref="GI:6018307"  
 /translation="MVSANTGTDDVDVHIDYLNLPATLEHIGITGPKAPQALVGM  
 VRAFTEAKLEVRLEERGPPWYKALVLYGSHVGLVGMPTDFRFSQGVHLMRIV  
 DGKIRDRDMPDFQGTILRQLGDPWDEGWRP"  
 7769..8821  
 /gene="snoGK"  
 7769..8821  
 /gene="snoGK"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative dTDP-glucose-4,6-dehydratase"  
 /protein\_id="AA01814.1"  
 /db\_xref="GI:6018308"  
 /translation="MASHTSATTDVILVTGAVGFIGSAYVRMLLENRAPGAGAPVR  
 VTVDLKITAGNLTNLDVARGDLRFVRGDIIDAEILDEMAHSDQVHFAESHSVDR

STRADADFLVTVQRLDAAALRHGVEPFLVSTDEVYSGSIAGSGWPBHPPLSPNS  
PYAASKADLMAFACHRTGLDVRVTRCSNNYPROHPKLIIRFVNTLLDGLFVPL  
YGDGNVRLWHLVHEDHCRGDLVTRAGPFGVYHIGGRSLNRELVMLELGCADW  
SSVRVPDRKCHDLRLSLDWGRARELGYRDPAREPSGLRSTVQWYRENRSWEPKLR  
GVTAPGTSVTPVGR"  
8857..9732  
/gene="snogC"  
8857..9732  
/gene="snogC"  
/codon\_start=1  
/transl\_table=11  
/product="putative dTDP-4-dehydrorhamnose reductase"  
/db\_xref="GI:6018309"  
/translation="MLASHLTAALAEFTGSRPAEAENVLGRALDITDGRAVDAAFA  
HPRVNVNCAFTDVGASRWAEAMRVNGSPRLARCAHGVRLHLHSTDIVFPG  
DTRSPYGESDAPRTVYGRSLAGERAVLSLLPDTGVTVRTVYVGGQGRSFVKTML

Query Match 27.0%; Score 135; DB 1; Length 16065;  
Best Local Similarity 61.5%; Pred. No. 5.4e-09;  
Matches 216; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
QY 150 GCGGGGAGTACAGATCAGTCTCGGACGCGCGGCGGAACTCTGCTTTCGCCAT 209  
DB 9743 GCGGGGATGGGACGATGAACGCGCGCGCTGTCTCCGGAAGCGCGTCTCTTCGAG 9802  
QY 210 GATCTCTGCGCGTGTTCAGCGGAAGACGAGCGGTAGTCCACCGCTCGGCGTGAA 269  
DB 9803 GATCTCGTGGCGTAGTTCAGCGGAGACACAGGTAGTTCGGGATGTCTTTCGCG 9862  
QY 270 CGCGTCCGGGTCGACCGGATGTCGTGCGGGGGTGAAGCGCGCTCTTTCGCGCG 329  
DB 9863 GTCTCGGAGCGTGCACCGGTATGTGGTTCGCGGACGACCGCTCCCTGCTCAGTTC 9922  
QY 330 CGTCGTTCGCACACCGGAGACGAGTCCGACCGATGCGCGAGGTTCGTACCGT 389  
DB 9923 GGTGGTTCGTCAGTATTCAGTCTTCAGCGGAGCGCGCGTTCGCGTTCGATGTGT 9982  
QY 390 GCGCGCTTTCGCGTTCGCGCGTACGCGACCGCTTCGCGCTTCGCGCTTCGCGAGTT 449  
DB 9983 GTTCCCTTCGCGGAGCAGCGTAGCGGATGCTTCGCGTTCGCGGAGGAGCG 10042  
QY 450 CAGCAGGCGAGCAGTTCGTCGGATGCTTCGCGTTCGCGGCGGCGAACT 500  
DB 10043 CAGCAGTTCGCGAGTTCGTCGGTTCGCGCGCGCTTCGCGTTCGCGGAGCT 10093

RESULT 16  
AX196110  
LOCUS AX196110 12152 bp DNA linear PAT 28-AUG-2001  
DEFINITION Sequence 182 from Patent WO0151639.  
ACCESSION AX196110  
VERSION AX196110.1 GI:15386342  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1  
AUTHORS Hosted, T.J., Horan, A.C. and Wang, T.X.  
TITLE Evernimycin biosynthetic genes  
JOURNAL Patent: WO 0151639-A 182 19-JUL-2001;  
Schering Corporation (US)  
FEATURES  
source  
1..12152  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

ORIGIN  
Query Match 26.2%; Score 131; DB 6; Length 12152;  
Best Local Similarity 100.0%; Pred. No. 1.9e-08;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACCCGACCGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGCGT 60  
DB 12022 GGTACCCGACCGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGCGT 12081  
QY 61 AGTCCGGGTAGACGGTGGCGGGAAGCGGTACCGCTTCGACGGTCAGCGGCGGCGG 120  
DB 12082 AGTCCGGGTAGACGGTGGCGGGAAGCGGTACCGCTTCGACGGTCAGCGGCGGCGG 12141  
QY 121 ACACCGCGCGG 131  
DB 12142 ACACCGCGCGG 12152

RESULT 17  
AX574140/c  
LOCUS AX574140 603 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 220 from Patent WO02079505.  
ACCESSION AX574140  
VERSION AX574140.1 GI:27551647  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1  
AUTHORS Farnet, C.M., Zazopoulos, E. and Staffa, A.  
TITLE Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
JOURNAL Patent: WO 02079505-A 220 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
source  
1..603  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
/note="var. africana"

## ORIGIN

Query Match 24.8%; Score 124; DB 6; Length 603;  
Best Local Similarity 100.0%; Pred. No. 3.6e-07;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGCGT 60  
DB 124 GGTACCCGACCGTGTCCCGGAAACAAGTTCGAGATACGGCGAGAGAACACCCCGCGT 65  
QY 61 AGTCCGGGTAGACGGTGGCGGGAAGCGGTACCGCTTCGACGGTCAGCGGCGGCGG 120  
DB 64 AGTCCGGGTAGACGGTGGCGGGAAGCGGTACCGCTTCGACGGTCAGCGGCGGCGG 5  
QY 121 ACAC 124  
DB 4 ACAC 1

RESULT 18  
AX574198/c  
LOCUS AX574198 15240 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 278 from Patent WO02079505.  
ACCESSION AX574198  
VERSION AX574198.1 GI:27551677  
KEYWORDS  
SOURCE Micromonospora carbonacea  
ORGANISM Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE 1  
AUTHORS Farnet, C.M., Zazopoulos, E. and Staffa, A.  
TITLE Compositions and methods for identifying and distinguishing  
orthosomycin biosynthetic loci  
JOURNAL Patent: WO 02079505-A 278 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)

```
FEATURES          Location/Qualifiers
  source          1..15240
                  /organism="Micromonospora carbonacea"
                  /mol_type="unassigned DNA"
                  /db_xref="taxon:47853"
                  /note="var. africana"

ORIGIN
Query Match      23.6%; Score 118.2; DB 6; Length 15240;
Best Local Similarity 53.3%; Pred. No. 8.3e-07;
Matches 249; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 33 AGATACGGGAGAGAGAACCCCGGCTAGTCGGGTAGACGCTGGCGGCGAAGGCGTAC 92
Db 12684 AGAGATGATCCAGGCGAGCAGCACATGCTCGGGACCGGGCTCCGGGTACGTCATGA 12625

QY 93 GCGCTTCGACGGTTCAGGGGCGGCGGACACCGGGGTCAGTCTACGCTGACGG 152
Db 12624 AAGTTGATCACTTCGGTGGGGCGGTCATCCCGTTCGGCGGACCGTCTCGACCG 12565

QY 153 GGGACGCTACAGGATCACTGTCGCGCAGCCCGCGGAACCTCTCTTCCTTCGCCATGAT 212
Db 12564 GGGCAGCGGAAGATCAACCGGGTCCGCTGGCCAGCATCTCCGCTCCCGGCGCAGAT 12505

QY 213 CTGCTCGGCGTGTCCAGGCGAGAGAGCGGCTAGTCCACCGGTCGGCGTGAACGC 272
Db 12504 CTCGTCGCGGAGTGCCAGGCGAGCAGGATAGTTCGGCGGCGCCCGCGACTC 12445

QY 273 GTCCGGGGTGCACACCGCGATGTCGTCGGGGGTGAGCGGCGCCCTCTGTCGCCCGCT 332
Db 12444 CTGCTCGTGTATTTTCGATGTCGTCGCGAGGTCGCGGCGCCACCTTGTCCGATT 12385

QY 333 CGTGTGCGACACCCAGAGACGAGTTCGGGACCGATCCCGAGAGTTCCTCAGGTGC 392
Db 12384 CGCTCTCGCGCGTACCGGATGAGCTCGCGTCTATCCGCGAAGTGCAGAGGGTGT 12325

QY 393 GCTCTTCGCTCGCGCGGTACCGACACCGCGTTCGCTCGCGCTTTCGCGAGTTTCAG 452
Db 12324 GCGCTTCGTCAGCGCGGTAGAGTGCACGCTCGCGGCGCTTGGCGCGAGTTCGCGCAG 12265

QY 453 CAGGCGAGAGGTCGTCGCGATGCGCTCGACGTCGCGCGCGGAACC 499
Db 12264 CAGGCGGCTACCTGTCACGCTGTCGCGGACCTGCTCGCGGAAGC 12218

RESULT 19
AY131229/c
LOCUS
DEFINITION
Streptomyces sp. 139 exopolysaccharide gene cluster, complete
sequence.
ACCESSION
AY131229 AF538923
VERSION
AY131229.2 GI:22657428
KEYWORDS
Streptomyces sp. 139
ORGANISM
Streptomyces sp. 139
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 31310)
AUTHORS
Wang, L., Li, S. and Li, Y.
TITLE
Identification and characterization of a new exopolysaccharide
biosynthesis gene cluster from Streptomyces
FEMS Microbiol. Lett. 220 (1), 21-27 (2003)
JOURNAL
MEDLINE
22531099
PUBMED
12644223
REFERENCE
2 (bases 1 to 31310)
AUTHORS
Wang, L., Li, S. and Li, Y.
TITLE
Isolation and sequencing of glycosyltransferase gene and
UDP-glucose dehydrogenase gene that are located on a gene cluster
involved in a new exopolysaccharide biosynthesis in Streptomyces
DNA Seq. 14 (2), 141-145 (2003)
3 (bases 1 to 31310)
JOURNAL
REFERENCE
Wang, L., Li, S. and Li, Y.
AUTHORS
Direct Submission
```

```
JOURNAL
Submitted (12-JUL-2002) Institute of Medicinal Biotechnology,
Chinese Academy of Medical Sciences, Tiantan Xili 1, Beijing,
Beijing 100050, People's Republic of China
4 (bases 1 to 31310)
AUTHORS
Wang, L., Li, S. and Li, Y.
TITLE
Direct Submission
JOURNAL
Submitted (03-SEP-2002) Institute of Medicinal Biotechnology,
Chinese Academy of Medical Sciences, Tiantan Xili 1, Beijing,
Beijing 100050, People's Republic of China
Sequence update by submitter
On Sep 3, 2002 this sequence version replaced gi:2252974.
FEATURES
  source
  1..31310
    /organism="Streptomyces sp. 139"
    /mol_type="genomic DNA"
    /strain="139"
    /db_xref="taxon:203783"
    /clone="cosmid pLV501"
  52..663
    /note="orf1; similar to Streptomyces coelicolor A3 (2)
    conserved hypothetical protein encoded by GenBank
    Accession Number AL512944 and to Streptomyces nodosus
    AmphORF1 encoded by GenBank Accession Number AF357202."
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="unknown"
    /protein_id="AA04226.1"
    /db_xref="GI:22657428"
    /translation="MAGSRVIVYAKDHLHGLRRDRREALLKLDGLSEYDARRPLT
    ATGNLLGLVATVEARYFYGFGRPSQPLRPWDADGSLWAAGGTRDGVVGF
    YRRWHSDDATDALDTPGRVPMPEPHADTNLFAVMVHLGETVHTGHADILRE
    GLDGTGLRPSHERPIDDEARAHAERAKLERAGAAESPLRSRS"
  660..1403
    /note="orf2; similar to Streptomyces coelicolor A3 (2)
    conserved hypothetical protein encoded by GenBank
    Accession Number AL158061 and to Bacillus halodurans
    BH0697 unknown conserved protein encoded by GenBank
    Accession Number AP001509; pfam03575; peptidase family
    S51."
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="unknown"
    /protein_id="AA04227.1"
    /db_xref="GI:22657430"
    /translation="MTSREPTIVATSGGHRVTRVNFDAVHVAVDLSGVHGRPR
    VNYGTATGDAEHFTARTEAARAGPDLTPLHLFPMNVDEATVLDQVWVWVG
    SVANLLAVRVHGLDGMRAWQAGVILSGVAGSLCWFGGATDSFQELRLTNAL
    GULPYGVHVDIDEGRRPLRHLVADGTLPTAHTDGGVGLVGTVEAVEAIEPG
    KGALVRRREGDEAVEERIEPRMLIVFSTIS"
  780..1173
    /note="similar to pfam03575 and peptidase family S51"
    /complement(1428..2192)
    /gene="stet"
    /complement(1428..2192)
    /gene="stet"
    /note="similar to gntR-family transcriptional regulators,
    including Streptomyces griseus Daser encoded by GenBank
    Accession Number AB061860 and Streptomyces coelicolor
    SC7E4.28 encoded by GenBank Accession Number AL359214;
    pfam00392"
    /codon_start=1
    /transl_table=11
    /product="putative gntR-family transcriptional regulator"
    /protein_id="AA04228.1"
    /db_xref="GI:22657431"
    /translation="MSTDVSSAGADTRAFGRDRVPKYLIKRLQMTDERAPGSPM
    PARELLAVEFGTSRTVRKALLELVSEGLRDIQKGTFRVAPKVTIQLTSTEDM
    RAQGLSPASQVLDIGVPADAEALAEVEPEDRVLRERLRAGPEMAIEATHLSA
    RRPFGRLNLTFTSLYTLASVGVGVRPAEDETTSATPREAGLLGTDLGLPMLL
    LSRHSDENGAPVWVRVSRGSRKYFARLRPRG"
    complement(1944..2114)
  misc_feature
  gene
  CDS
```

```

/misc_feature
/notes="Region: pfam00392 and smart00345; similar to gntR
family and helix-turn-helix gluconate operon
transcriptional repressor"
2530..3843
/gene="ste2"
2530..3843
/gene="ste2"
/notes="similar to putative sugar transporter sugar-binding
proteins, including Streptomyces coelicolor SC7E4.29
encoded by GenBank Accession Number AL359214 and
Streptomyces griseus Dasa encoded by GenBank Accession
Number AB061860; pfam01547; PS00013; contains possible
N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative sugar transporter sugar-binding
protein"
/protein_id="AA04229.1"
/db_xref="GI:22657432"
/translations="MGAQGIPEEGWVKEKLAIAIGVAGMLFSAACGGDDGGKKA
GUDYAGQILTVWMDGSSPDQKDLTAFEFKTKYAKVKEVQQRNGIQKLTALS
ENPPDVFIGNQTAYAKTGLADLADLKEIGWETESLNESSVFEKQYAPWY
FANRVLYNKKWADAGLDKPTKDRDEFYALKTIKAKTDAPFIYFGQNVHFVGLV
VGGSLVKGDKYVNLNDDPKVAKAETTKFQALSKAPKDEATPQQGEVFAKQ
NVGAFICMGWEAGIAIEANPAIEKEIGYFTIPGATADKEPVFLGSLNLAAGSQKQ
ELAKFLKALSPLKSGQLKNGVTPNKESIQGNLKNAPAEAPAAAVAGGTPLI
PEWAAVENDPNPVKTYLTAVLNGKSPAAAKVGESEFNKRLAQQQ"
2560..2595
/misc_feature
/gene="ste2"
/notes="Region: PS00013; similar to prokaryotic membrane
lipoprotein lipid attachment site; positioned
appropriately within protein"
2592..3762
/misc_feature
/gene="ste2"
/notes="Region: pfam01547; similar to bacterial
extracellular solute-binding protein"
3932..4918
/gene="ste3"
3932..4918
/gene="ste3"
/notes="similar to putative ABC transporter integral
membrane proteins, including Streptomyces coelicolor
SC7E4.30 encoded by GenBank Accession Number AL359214 and
Streptomyces griseus DasB encoded by GenBank Accession
Number AB061860; pfam00528"
/codon_start=1
/transl_table=11
/product="putative sugar transporter integral membrane
protein"
/protein_id="AA04230.1"
/db_xref="GI:22657433"
/translations="MTVQTERPPSGADVVRKSGGTAPGPPPPRAARAGALAPYLILL
PACATVLLGWPLLDLLSDFONMAQLIQHVTENGDYKKEVLGTEDFWRVTLR
SILFTAVNALVLLGLALVGLLARKMRRLTMIGLVWAMPVVAATTVYQWLFA
QRGVNVLDDLGWTSMDYSWTSSCMSTFFVVTLLIVMVSVPFVIALYAAATTP
GELYEAALDAGANWSEFTVTLPELRPELYATELEILWIKAFQVTFEGGPDPR
LTSLPVIAYIEGVGHQVHGMAAIVLILMLGLTATYLRILVKAQEDAL"
4577..4774
/misc_feature
/gene="ste3"
/notes="Region: pfam00528; similar to
binding-protein-dependent transport systems inner membrane
component"
4915..5748
/gene="ste4"
4915..5748
/gene="ste4"
/notes="similar to putative ABC transporter integral
membrane proteins, including Streptomyces coelicolor
SC7E4.31 encoded by GenBank Accession Number AL359214 and
Streptomyces griseus DasC encoded by GenBank Accession
Number AB061860; pfam00528; PS00013; contains possible
hydrophobic membrane-spanning domains"

```

```

/codon_start=1
/transl_table=11
/product="putative sugar transporter integral membrane
protein"
/protein_id="AA04231.1"
/db_xref="GI:22657434"
/translations="MKESLGRVMPNATALVVFVACLPVYVMPATALKPTGDIASD
PVWPTVTEHEPATADADHFYTYVNSLTVCCLAVLFSGLIALAGSPALARMFKG
RGRFVGLMAQMAPNEVWVIALYIMIVDRSMNLNLSVLTAFYMMMLPPTLILRGP
VAAVPKLESAMVDGTRGOAFRRVILPLLAGLMSFTMGFPTAWNEFPVLVLVSK
ESGGAQTLPLWLSSFTQFGDDWGATMAASSLFAVPTLILFVLQRAVSGLTAGAVK
G"
4948..4983
/misc_feature
/gene="ste4"
/notes="Region: PS00013; similar to prokaryotic membrane
lipoprotein lipid attachment site"
5320..5616
/misc_feature
/gene="ste4"
/notes="Region: pfam00528; similar to
binding-protein-dependent transport systems inner membrane
component"
5674..8152
/gene="ste5"
/notes="steA"
5674..8152
/gene="ste5"
/notes="steA; possible glycosyl transferase; similar to
glycosyl transferases, including Streptomyces coelicolor"

```

```

Query Match      23.2%; Score 116.2; DB 1; Length 31310;
Best Local Similarity 58.6%; Pred. No. 1.2e-06;
Matches 202; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
QY 154 GGGAGGTACAGGATCCACTGTCCCGCCAGCCGCGGGAACCTCCTCTCTCTCCCATGATC 213
DB 25124 GGGATGGGAAGACAGCCGCCGCCCTCTCGTGCAGGAGACAGTCTCGACCAAGC 25065
QY 214 TCGTCGGCGTGTTCAGGCGAAGAGACGCCGTAGTCCACCGCTCGGGCGTGAACGCG 273
DB 25064 TCGGCCCGCAGGTTCCAGGGCAGCACCCAGACAGTAGTCCGCGCGGCGCATCCGG 25005
QY 274 TCGGGGGTGGCCACCGGATGTGGTCCCGGGGTGAGCCCGCCCTGTCTCGCGCGGCTC 333
DB 25004 TCGGGCGCAGATCGGATCGGGTGCCTCGGGGTGTACCGCGCGTCTCTAGGGTTG 24945
QY 334 GTGTGCGACACCCAGGAGACCCAGGTTCGGACCGATGCGCGAAGAGTTCTGTCACGGTGGCG 393
DB 24944 CGGTCCACCGTGTACGGCAGCAGGTTCGGCGCGCGACCGCGCAGTGTGTTGAGCAGGTTGTG 24885
QY 394 CTCCTTCGCGGTGCGCGCGTACGCCACCCCGCTTGCCTTCGCGCTTGAGCAGTTACAG 453
DB 24884 CCTTTCGCGCGCGCGCTTACCGACGACCGTCTTCTCGCTCTCGCGCGTTCGATCAGG 24825
QY 454 AGGGCGACAGGTGGTTCGGATGCCCTCGACGTCCGCGCGGGAAC 498
DB 24824 AACCGCAGAGTCCCGCGCACCTTGGCCACCGCGCGGCGGAGAAC 24780

```

## RESULT 20

AX195978/c

LOCUS

Sequence 50 from Patent WO0151639.

DEFINITION

AX195978

VERSION

AX195978.1 GI:15386210

KEYWORDS

SOURCE

Micromonospora carbonacea

ORGANISM

Micromonospora carbonacea

Bacteria; Actinobacteria;

Actinobacteridae; Actinomycetales;

Micromonosporaceae; Micromonospora.

REFERENCE

1

AUTHORS

Hosted, T.J., Horan, A.C. and Wang, T.X.

Evenimycin biosynthetic genes

Patent: WO 0151639-A 50 19-JUL-2001;



FEATURES Schering Corporation (US) Location/Qualifiers  
source 1..1236  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
1..1236  
CDS  
/note="unnamed protein product; evrM"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC60015.1"  
/db\_xref="GI:15386211"  
/translation="MVERITTCRVGGTGLRVLGLAQYLGGSFVKGVSDPRVRPPL  
ELRCARDGGDPCGLVQLRHPTLGLLVDYWRINDTWRHLADYVDAEVLG  
ERPKVLIDICGNLTLGLNLTDAQRWIDPSSAADAPDFNLIDFFPSPELVPHAG  
SEPLATSIAMEYDDEVEFARQVEALLAPGVVVEVYLPAMLEIAGVDSITCHEHL  
SVYSLATLSILGALGLEIYRASINGVNGSICCVTVRAGEVGGADGSKALADREL  
ALGLDAPQRFARQVQRDRDEVSALLRELRGQETVHVYGSTGNTLLQFCGIDR  
ELRIYAERNPDKVARTLGTDLIIISQESRAARPDYLLVFWHFRDIVEAREBML  
ASGRLFIPLPRVETVGRPG"

Query Match 22.6%; Score 112.8; DB 6; Length 1236;  
Best Local Similarity 56.5%; Pred. No. 8.4e-06;  
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;  
ORIGIN  
QY 128 CGGTCACGTCGTACGTCGTCGCGGGGAGCTACAGATCCACTGTCCGCGCAGCCCGC 187  
Db 1276 CGGTCGCGCCACCGTCTCGACCGCGGGCAGCGGAAAGATCAACCGGTGCGGTGGCCA 1217  
QY 188 GGAATCTCTGCTCTTCCCTGCCATGATCTCTGTCGGCGTGGTTCAGCGGAGAGCAGCGGT 247  
Db 1216 GCATCTCCGCTCCCGGCGCAGATCTCTGTCGGCGGAGTGCAGGCGCAGGACGAGTAGT 1157  
QY 248 AGTCACCGCGTCGCGGTCGACCGGTCGCGGGTGGCGACCGGATGTCGTCGCGGGGG 307  
Db 1156 AGTCGCGGGCGGCGCGCGGACTCTCTGTCGTCGATGATTTGATGTCGTCGCGAGG 1097  
QY 308 TGAGCGCGCCCTGCTTGGCGCGGTCGTCGTCGACACCCAGGAGACAGGTCCGACCGA 367  
Db 1096 TCGCGCGCCACCTGTGTCGCGATTCGCGCGGCTACCGGATGAGCTCGCGGTCTA 1037  
QY 368 TCGCGCAGAGTTCGTCAGGTCGCGGTCGTCGCGGTCGCGCGGTCGCGGTCGCGGTCG 427  
Db 1036 TCGCGCAGAACGTCAGGAGGTTGTCGTCGTCGCGGTCGCGGTCGCGGTCGCGGTCG 977  
QY 428 TCGCTTCGCGCTTGAGCGAGTTCAGCGGCGAGCAGTCGTCGCGGATGTCGTCGAGCT 487  
Db 976 GCGCTGCGCGCGCGCGGCTCGCGCAGCGGCGCTCACCTCGTCAGGTCGTCGCGCACCT 917  
QY 488 CGCGCGCGAACC 499  
Db 916 GCTCGCGGAGC 905

RESULT 22  
SEERYABIO 4254 bp DNA linear BCT 30-JUN-1993  
LOCUS S.erythraea erythromycin A biosynthesis gene cluster encoding  
DEFINITION hydroxylase; thioesterase; methylase.  
X60379.1 S69652  
ACCESSION X60379.1 GI:48941  
VERSION erythromycin biosynthesis; hydroxylase; methylase; thioesterase.  
KEYWORDS erythromycin biosynthesis; hydroxylase; methylase; thioesterase.  
SOURCE Saccharopolyspora erythraea  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Pseudonocardineae; Pseudonocardaceae; Saccharopolyspora.  
REFERENCE 1 (bases 1 to 4254)  
AUTHORS Haydock, S.F., Dowson, J.A., Dhillon, N., Roberts, G.A., Cortes, J. and  
Leadlay, P.F.  
TITLE Cloning and sequence analysis of genes involved in erythromycin  
biosynthesis in Saccharopolyspora erythraea; sequence similarities  
between Eryg and a family of S-adenosylmethionine-dependent  
methyltransferases  
JOURNAL Mol. Gen. Genet. 230 (1-2), 120-128 (1991)  
MEDLINE 92079886  
PUBMED 1840640  
REFERENCE 2 (bases 1 to 4254)  
AUTHORS Haydock, S.F.  
TITLE Direct Submission  
JOURNAL Submitted (19-AUG-1991) S.F. Haydock, University of Cambridge,  
Tennis Court Road, Cambridge CB2 1QW, UK  
FEATURES  
source 1..4254  
/organism="Saccharopolyspora erythraea"  
/mol\_type="genomic DNA"  
/strain="NRL2338"

FEATURES Schering Corporation (US) Location/Qualifiers  
source 1..1236  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
1..1236  
CDS  
/note="unnamed protein product; evrM"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC60015.1"  
/db\_xref="GI:15386211"  
/translation="MVERITTCRVGGTGLRVLGLAQYLGGSFVKGVSDPRVRPPL  
ELRCARDGGDPCGLVQLRHPTLGLLVDYWRINDTWRHLADYVDAEVLG  
ERPKVLIDICGNLTLGLNLTDAQRWIDPSSAADAPDFNLIDFFPSPELVPHAG  
SEPLATSIAMEYDDEVEFARQVEALLAPGVVVEVYLPAMLEIAGVDSITCHEHL  
SVYSLATLSILGALGLEIYRASINGVNGSICCVTVRAGEVGGADGSKALADREL  
ALGLDAPQRFARQVQRDRDEVSALLRELRGQETVHVYGSTGNTLLQFCGIDR  
ELRIYAERNPDKVARTLGTDLIIISQESRAARPDYLLVFWHFRDIVEAREBML  
ASGRLFIPLPRVETVGRPG"

Query Match 22.6%; Score 112.8; DB 6; Length 1236;  
Best Local Similarity 56.5%; Pred. No. 8.4e-06;  
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;  
ORIGIN  
QY 128 CGGTCACGTCGTACGTCGTCGCGGGGAGCTACAGATCCACTGTCCGCGCAGCCCGC 187  
Db 1234 CGGTCGCGCCACCGTCTCGACCGCGGGCAGCGGAAAGATCAACCGGTGCGGTGGCCA 1175  
QY 188 GGAATCTCTGCTCTTCCCTGCCATGATCTCTGTCGGCGTGGTTCAGCGGAGAGCAGCGGT 247  
Db 1174 GCATCTCCGCTCCCGGCGCAGATCTCTGTCGGAGTGCAGGCGCAGGACGAGTAGT 1115  
QY 248 AGTCACCGCGTCGCGGTCGACCGGTCGCGGGTGGCGACCGGATGTCGTCGCGGGGG 307  
Db 1114 AGTCGCGGGCGCGCGCGGACTCTCTGTCGTCGATGATTTGATGTCGTCGCGAGG 1055  
QY 308 TGAGCGCGCCCTGCTTGGCGCGGTCGTCGTCGACACCCAGGAGACAGGTCCGACCGA 367  
Db 1054 TCGCGCGCGCCACCTGTGTCGCGATTCGCTCGCGGCGTACCGGATGAGCTCGCGGTCTA 995  
QY 368 TCGCGCAGAGTTCGTCAGGTCGCGGTCGTCGCGGTCGCGGTCGCGGTCGCGGTCGCG 427  
Db 994 TCGCGCAGAACTGACGAGGTTGTCGCTTCGTCGCGGCGGTCGCGGTCGCGGTCGCG 935  
QY 428 TCGCTTCGCGCTTGAGCGAGTTCAGCGGCGAGCAGTCGTCGCGGATGTCGTCGAGCT 487  
Db 934 GCGCTGCGCGCGCGAGCTCGCGCAGCGGCGCTCACCTCGTCAGGTCGTCGCGCACCT 875  
QY 488 CGCGCGCGAACC 499  
Db 874 GCTCGCGGAGC 863

RESULT 21  
AX574042/c 1281 bp DNA linear PAT 07-JAN-2003  
LOCUS Sequence 122 from Patent WO02079505.  
DEFINITION AX574042  
ACCESSION AX574042  
VERSION AX574042.1 GI:27551598  
KEYWORDS Micromonospora carbonacea  
SOURCE Micromonospora carbonacea  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE 1  
AUTHORS Farnet, C.M., Zazopoulos E. and Staffa, A.  
TITLE Compositions and methods for identifying and distinguishing  
orthomycin biosynthetic loci  
JOURNAL Patent: WO 02079505-A 122 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)



```
/db_xref="taxon:1836"
69..1313
/gene="orf1"
CDS
69..1313
/gene="orf1"
/codon_start=1
/transl_table=11
/protein_id="CAA42926.1"
/db_xref="GI:48942"
/db_xref="SPTREMBL:Q00463"
/translation="MIFLVGLCKRICGNHDLSEVLHLDQALTGVPFRNDEIVPSV
PLELVKCAPGGGLVQLRESADPGIMYEGVYSGIRPFNHLGKVKALRGWVPV
GPDLDVDDIGSLLRGKLPDAKLAGFDLVGKPRDLVPEPADLVTFPSADAFE
EYGERAKVTSIAMPVDLPMPFRMDVHDLDADGLWLMQSYLPSMDAGADIV
VCHLELYGLAQIEWMAQRVGLKVDAEITDVYGGSRVAVLPKGGSHPVDEAGLR
IARERAAKLDTMAYEPARETERQDQLLEFLAKSRABGKLTIGYGASTKGNVILQ
YGLTEQDLPICGEVSPKSGCYTPTRIPIVSEBEAKSRPDQLLVLPFIYRDGFEV
RQEFLAGGCKLIIFLPLEVY"
1310..2530
/gene="orf2"
1310..2530
/gene="orf2"
/codon_start=1
/transl_table=11
/product="hydroxylase"
/protein_id="CAA42927.1"
/db_xref="GI:48943"
/db_xref="GOA:Q00441"
/db_xref="SWISS-PROT:Q00441"
/translation="MTVPLESDSFHVDVWYTYAELRETAPVTVFVFLGQDAMLVTG
YDEAKALSDRLSDPKKVPQVEVPAYLGFPPEDVNTFATNMGSDPPTHRLR
KLVSQETFLVREAMPKVEQVETAELLDEVDGSGVVDIVDFRFAHLPKIVICELGVD
KYRGFGFSGSEILVMDPERAEORGQAAREVNFILDLVERRTEPGDGLLSALIRV
QDDGGLSADSLTALVLLAGFEASVSLIGICTYLLLTHTPDQDLAVRBPDSAL
PNAVEILRYIAPPETTPFAAEVEIRGVAIPOYSTVLVANGAANRPKFPDPDRFR
DVTRGDHGLSGGQHFCWGRPLKLEGEVALRFGFRFALSIGDADDVVMRRSL
LNRGDHLPVRLDG"
2527..3270
/gene="orf3"
2527..3270
/gene="orf3"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="CAA42928.1"
/db_xref="GI:48944"
/db_xref="GOA:Q00442"
/db_xref="SPTREMBL:Q00442"
/translation="MSTWLRREGPPVEHRLVCPFHAGAAADSVLDLARALAPIDV
DAVQPGQRDRDEEPLGTAGEIADEVAALVRASGGDGFALFGHSMGALIAETARR
LEKEPGGGLRUFVSGQTAPRVHERNTDLPDGLVDLRLRGTSEALADBALLAMS
LPVLRADYRLRSYAWADGPPLRAGITLPCGADPLTATGDAERLQHSVIPGRTTF
PGGHFYLGQVTEVAGAVRDILIRAGLAG"
complement(3247..4167)
/notes="orf4"
/codon_start=1
/transl_table=11
/product="methylase"
/protein_id="CAA42929.1"
/db_xref="GI:581650"
/db_xref="GOA:Q54095"
/db_xref="SPTREMBL:Q54095"
/translation="MSVKQKSLQDLVDFAKWHVTRVPPSSRLAYALPADDEAT
TEGAINVGLGKPCAGLEANEQNLQAEAGQSEGEVDLDVGFGLGAQDFWLET
RKPARIVGDLTSPSHVRIASERARENQDRLQFKEGSATDLPFGARTFDRVTSLESA
LHNPVETDFPKAFVLPKGGVLAIGDIIPLDLRFPGSDGPKLAPORSGLSGIPV
ENWVPRETKALQREAGFVDEVKVDNVMPEWFLDYMLRKLQDESFKKSVSRLPYSQ
VKRSLTSDSGMKGLPDLDFVIASARKPGA"

gene
21.8%; Score 108.2; DB 1; Length 4254;
Best Local Similarity 55.8%; Pred. No. 2.4e-05;
Matches 206; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

ORIGIN
```

```
Qy 130 GGTGAGTCTGCTACGTGTACGCGGGGACGTACAGGATCCACTGTCTCGCCAGCCCGCGG 189
|||
Db 1318 GGTGCTCATACGACTTCCAGTCGGGTAGGGGGAAGATCAGCTTCCGCGCCAGCGAGG 1259
|||
Qy 190 AACTCTGCTCCTTCGCCATGATCTGTCGGCGTGTTCAGCGGAGAGAGCGGCGTAG 249
|||
Db 1258 AACTCTGCTCCTTCGCGTTTCGAGAACCCGTCGCGGTAGATCCAGGGGACACAGAGCTGG 1199
|||
Qy 250 TCACACCGCTCGGGCGGTGAACGGGTCCGCGGATCGCACCGGATGTGCTGCGCGGGGCTG 309
|||
Db 1198 TCGGGCGCGCGACTTGGCTCTTCTCCGACACAGATCGGGTCCGGGTGCGGGGGTG 1139
|||
Qy 310 AGCGGGCCCTGCTTGGCGCGGTGCTGTCGACACCCAGGAGACAGGTCCGACCGATG 369
|||
Db 1138 TAGCAACCGGACTTCTCCGGGCTGACCTCCCGATGACGAGGCGAGGTCTCTCGGTGAGC 1079
|||
Qy 370 CCGCAGAAAGTTGCTACGCTGGGCTCTTCGCGCTCGCGCGATGACGACACACCGCGCTG 429
|||
Db 1078 CCGCAGTACTGCAGGATGACGTTGCCCTTGTGCGAGGCGCGTACGCGAGGTCAGCTTG 1019
|||
Qy 430 CCTTCGCGCTTACGAGGATTCAGCAGGGCGAGCAGGTGCGTTCGAGATGCCCTCGACGTCG 489
|||
Db 1018 CCTTCGCGCGGGGACTTGGCAGGAACTCCAGCAGTTGTCGCGCTCGCGCTCGGTTCG 959
|||
Qy 490 GCGCGGAAC 498
|||
Db 958 GCGCGGAAC 950
|||
```

```
RESULT 23
RME603644/c
LOCUS Rhizobium meliloti 315000 bp DNA linear BCT 05-JUL-2002
DEFINITION Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete pSymB;
segment 3/6.
ACCESSION AL603644 AL591985
VERSION AL603644.1 GI:15140405
KEYWORDS
SOURCE Sinorhizobium meliloti (Rhizobium meliloti)
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
REFERENCE
1 Finan,T.M., Weidner,S., Wong,K., Buhrmester,J., Chain,P., Gouzy,J.,
Vorholter,F.J., Hernandez-Lucas,I., Becker,A., Cowie,A., Gouzy,J.,
Golding,B. and Puhler,A.
The complete sequence of the 1,683-kb pSymB megaplasmid from the
N2-fixing endosymbiont Sinorhizobium meliloti
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)
21396508
MEDLINE 11481431
PUBMED
REMARK
REFERENCE 2 (bases 1 to 315000)
AUTHORS Weidner,S.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld,
Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld,
Germany
COMMENT Submitted on behalf of Universitaet Bielefeld, Biologie IV
(Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and
Department of Biology, McMaster University, 1280 Main Street West,
Hamilton, Ontario, L8S 4K1 Canada
mailto:Stefan.Weidner@genetik.uni-bielefeld.de
PEXO, pSymB.
FEATURES
source Location/Qualifiers
1..315000
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymB"
complement(52..405)
/gene="SMB20542"
gene
```

```

CDS
complement(52..405)
/gene="SMB20542"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative protein in ISRM14"
/protein_id="CAC48921.1"
/db_xref="GI:15140406"
/db_xref="GOA:Q9X6A1"
/db_xref="SPTREMBL:Q9X6A1"
/translation="MIVAGORLPILIAITRPVDRCGHOALAMVOTELKLDPHSGVTV
IFRSKGRDRIKILVDTGTMVLTIKILEHSGPAWPKYQDGTWLSRGOYBALFEGLDW
RVNAQRVTPSAAG"
289..705
/gene="SMB20544"
289..705
/gene="SMB20544"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative protein in ISRM14"
/protein_id="CAC48922.1"
/db_xref="GI:15140407"
/db_xref="GOA:Q9X6A0"
/db_xref="SPTREMBL:Q9X6A0"
/translation="MRVELQLQHLHQSLRMPTAEVHRPCNDRQSLARDHDTGARM
ARTAAATSTGTGPTRTNTSGPISNVKTDETASDGGAKTFTASLGSAAIAKGTNAG
SASPLSDRPGIASNGARMPSRACKRQKESNCGTTS"
complement(402..845)
/gene="SMB20543"
complement(402..845)
/gene="SMB20543"
/function="ELEMENTS OF EXTERNAL ORIGIN; Transposon-related
functions"
/note="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative protein in ISRM14"
/protein_id="CAC48923.1"
/db_xref="GI:15140408"
/db_xref="GOA:Q92W24"
/db_xref="SPTREMBL:Q92W24"
/translation="WAGDGFVGRYVVEPRGNRWDPAVKARIVASSLEPGYVVDV
ARRHDVPHQUSFWRRQAREGILALPFEMFGLSESGDAEPFVPLAIAEPSEAVNV
LAPPISEAVSVLTLEIGFDVVVPGDVVERVAALVRAMRPV"
complement(1438..1656)
/gene="SMB20545"
complement(1438..1656)
/gene="SMB20545"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/note="Product confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAC48924.1"
/db_xref="GI:15140409"
/db_xref="GOA:Q92W23"
/db_xref="SPTREMBL:Q92W23"
/translation="MDDRAGSAADAEPAPDISVAQCAAAADLLDFAAQLAGLAEDL
KALAARPIPTPTLAEPPHPEATSSDL"
complement(1843..2361)
gene

CDS
complement(1843..2361)
/gene="SMB20546"
/function="MISCELLANEOUS; Hypothetical/Global homology"
/note="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHEITICAL PROTEIN"
/protein_id="CAC48925.1"
/db_xref="GI:15140410"
/db_xref="GOA:Q92W22"
/db_xref="SPTREMBL:Q92W22"
/translation="MAERWDECVIIDLENLEGIQIVMSPAYAAQLLENWPDPTGRSY
SAAALNSCTDMFGAASAAAPARDAFLAAVEARVKTLRWHHAAPLVKGNDSIEGGYVA
ELGFOARSHSVRAWSTGGGQADREFSARSSRSSESGAVLVSPSTDARLISFTLPPA
ASRTCAVSAHW"
complement(2518..2796)
/gene="SMB20547"
complement(2518..2796)
/gene="SMB20547"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/note="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAC48926.1"
/db_xref="GI:15140411"
/db_xref="GOA:Q92W21"
/db_xref="SPTREMBL:Q92W21"
/translation="MYEPFAQIKAWLKEGFTLSAAGVLRRLTEIDPIRFTKNIWMV
QMLVKAWRMEMAVLVLDGWTKSLPSAAAAADVTGTSASAVLGNLR"
3930..4424
/gene="SMB20548"
3930..4424
/gene="SMB20548"
/function="MISCELLANEOUS; Unknown"
/note="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHEITICAL PROTEIN"
/protein_id="CAC48927.1"
/db_xref="GI:15140412"
/db_xref="GOA:Q92W20"
/db_xref="SPTREMBL:Q92W20"
/translation="MESIIERNCRKQDFDIFMPSFYTELHRTKQILQKRPFPLVGY
GFVSLPRLNFEELRRVDGVVCFPLRGANYPLEFPDGYHGPVLRSRATAGLPQAAAL
PEGDRAPRANPALGRPADPAKGSERARLHGRPCRCYRFSKPADQASADNYQWIE
QPHC"
6048..6365
/gene="SMB20549"
6048..6365
/gene="SMB20549"
/function="MISCELLANEOUS; Hypothetical/global homology"
/note="Product confidence : hypothetical"
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHEITICAL PROTEIN"
/protein_id="CAC48928.1"
/db_xref="GI:15140413"
/db_xref="GOA:Q92W19"
/db_xref="SPTREMBL:Q92W19"
/translation="MRALKASAHYDVMAEQIRALPVQOLASMNCLIVSRATAPHLA
TIVCECLKWGFYKSPMAWKTKTAAGKVRNGTGRVTRTTGIVFGVLTGLNGPKSHVFP
TIS"
complement(6533..6778)
gene
```

```
/gene="SMD20550"
complement(6533..6778)
/gene="SMD20550"
/function="MISCELLANEOUS; Hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC48929.1"
/db_xref="GI:15140414"
/db_xref="GOA:Q92W18"

Query Match
Best Local Similarity 21.1%; Score 105.6; DB 1; Length 315000;
Matches 195; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 154 GGCAGCTACAGATCCACTGTCGCCAGCCCGCGGAATCTCTCTCTCCATGATC 213
Dd 125772 GGCACAGGAATGATAAATTCACACCGCTGTGATGTCGGCTTGTGGCGGAGATC 125713
QY 214 TCCTCGCGTGTTCAGGAGAGACGCGGTAGTCCACCGCTCGCGGTGACCGG 273
Dd 125712 TCCTCGCGGAAATTCAGGAGAGACGCGGTAGTCCCGGTGCGCTTGGACAGGGCC 125653
QY 274 TCCTCGCGGTGCGCACCGCGATGTCGTCGCGGGGTGAGCGCGCTTGTGCGCGCGTC 333
Dd 125652 TCCTCGCGGTATCGCGAGAGCTGCGCGAGCATATCTTGCCTTGTGCGCGTGTG 125593
QY 334 GTCTCGCACCCAGAGACAGGTCGCGACCGATCCGACGATTCCTGTCAGGTGGCG 393
Dd 125592 CGTCTCGACGAGGAGTCTTAGAGTTCACGTCGATCCGATCCGATCGTGTGATCAGCGTCGT 125533
QY 394 CTCTTCGCGCTCGCGCGTACGCGACACCGCTTCCCTCGCGCTTGGCGAGTTCAGC 453
Dd 125532 CCCTTGGCGCGCGCCCAAGCGGCAATCGAGGAGCCCTTGATTCAGCCTGTGCGC 125473
QY 454 AGGCGAGAGGTCGTCGCGGATGTCGCTCGAGTCGCGCGCGCAA 497
Dd 125472 AGTGCGAGAGCTCACGCTTCAATACGTCGACCTTCGACAGAGAA 125429

RESULT 24
AF147704
LOCUS
DEFINITION
Streptomyces fradiae NDP-hexose 2,3-enoyl reductase TylCII
(tylCII), NDP-hexose 4-ketoreductase TylCIV (tylCIV), NDP-hexose
3-C-methyltransferase TylCIII (tylCIII), mycarosyl transferase
TylCV (tylCV), and NDP-hexose 3,5- (or5-) epimerase TylCVII
(tylCVII) genes, complete cds.
AF147704
AF147704.1 GI:5305790
Streptomyces fradiae
Streptomyces fradiae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 5908)
Bate.N., Butler,A.R., Smith,I.P. and Cundliffe,B.
The mycarose-biosynthetic genes of Streptomyces fradiae, producer
of tylosin
Microbiology 146 (Pt 1), 139-146 (2000)
20121747
10658660
REFERENCE
2 (bases 1 to 5908)
Bate.N. and Cundliffe,B.
Direct Submission
Submitted (30-APR-1999) Biochemistry, University Of Leicester,
University Road, Leicester LE1 7RH, UK
Location/Qualifiers
1..5908
/organism="Streptomyces fradiae"

/mol_type="genomic DNA"
/strain="759235"
/db_xref="taxon:1906"
complement(80..1069)
/gene="tylCII"
/notes="orf10*"
/complement(80..1069)
/gene="tylCII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 2,3-enoyl reductase TylCII"
/protein_id="AAD41821.1"
/db_xref="GI:5305791"
/translation="MSCMYVOLGRGATLVSRMLGTVNPSGRVDPDDAIRLMDALDR
GVNCDVADIRYGLRWLGRGRRDDVLAATKVGSEPSDRVNDGLSA
RHVIRSCAELGHDHIDLYQHRMDRTVRWDELQWMDQLVASKRVYIGSSNFA
GWHLAQGEAARGSLGSEOCNLVLAHAEVLVPAARAYGIGVFAWSPHGGGL
LSGALRLAGSTAVKSCQGAORTLPALRTIARYERFCARVGRDPAEVLGMLLSP
GVSGAVTGPRTGHLVSAFVLELESEEHEHLEALFPFVSGSGVEPEAQN"
complement(1066..2109)
/gene="tylCIV"
/notes="orf9*"
/complement(1066..2109)
/gene="tylCIV"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 4-ketoreductase TylCIV"
/protein_id="AAD41822.1"
/db_xref="GI:5305792"
/translation="MSMPLVPGPARDRRLVLLGASGFLGSAVAELARRPVRLRL
VAREPSPVAGAAVEVRBDLSLPEVTEAEDADAVVLAHTGGEKSWAAGER
SEHNVGLMRELAELRPTGPADATRAPVLFSTLQAGMEQATFGTYAAQKLAER
VHNRADAGRVAGVLLTVIGRSPITGSPRGVIAVTRAGALAGDPIFTHWDGSEV
RDLVDADAATAFTALEHADQLRGKHWVGTGRHRLDRVFGTVLAALAAHGTGRPE
PVTVDPGVAEFTCDTPDPSAFRAVNTGMRPRAEPADGIAAATAAAGAGDSPEP
EGAKRGCGTERR"
complement(2106..3362)
/gene="tylCIII"
/notes="orf8*"
/complement(2106..3362)
/gene="tylCIII"
/function="biosynthesis of NDP-mycarose during tylosin
production"
/codon_start=1
/transl_table=11
/product="NDP-hexose 3-C-methyltransferase TylCIII"
/protein_id="AAD41823.1"
/db_xref="GI:5305793"
/translation="MPAVPREDMIIISACVCGNRELLPVLDELGEQALGVFPRTREE
TVPSIPLGLVKSPQCGVLQRLRTPDGLMYGEGYGRSGIRPFMHLRRVAAIR
ELVDLGPDDLVDIGSNDATLKAYPEGRLVIGIDPSGDKRFLYPPHAEILYFSR
DVTAREGTRARVITSIAMFYDLPDPLAFMRDHDVLDADDGVMWEOQSLPAMLEAD
AYLVCHHELYVALQCIEMAEAGLTVLRLDTVYCGSLCVTLARASPHRDEA
GPARKAREAKINTMAPFEFARVEHORDALRDLFRDRAAGRLTLTGAGSTKGN
VILQYCGIGERDLPCIGVSEKAGRTTGTGTGIPVSEEDAKAMPDQLLVPIWYRE
GVFEREDFLAGGRLVFPPLRLDVV"
3575..4960
/gene="tylCV"
/notes="orf7*"
/complement(4960)
/gene="tylCV"
/function="catalyses the incorporation of mycarose during
tylosin production"
/codon_start=1
/transl_table=11
/product="mycarosyl transferase TylCV"
/protein_id="AAD41824.1"
/db_xref="GI:5305794"
```

/translation="MAGLRPGAGVPPGCTPMPISPGKHCVRSATSRRLRPLYPHAGTPL  
GRERRADEGSCWAHIAFILLPAAGHVNPTLGAABELAARGHVTVALPEDMDARAV  
RVGATVDFDRERFADNVPEKESDVTDEGEFLKYLEWLLDTTADTFLLESFAA  
EDPDVANDPSTFWKLLAGKWDIPVIRKSTFASNEHWAHPFPEFGACQVDFPA  
IEFARAKLLKEHGTSDVFAATVQSGGLFYMPRYFQVAGETFDHRRHFVGGCA  
PRAFHGTWQRPEDRGHTVPPGQKIVTCVRGAARTWVLDLVGSPFTGGHVVVG  
GDAGPLPENLVDRFVPLGVDHTLLVNHGCTAMEALAHGVPIVAMPPEPRA  
TARRIABLDLGDWLLPGEVTAELKSLGIAQRLVLTDRIRKGLDRMRGEIRRAGGPAVA  
DVIEGLSPAA"  
5004. 5597  
/gene="tylcvii"  
/note="orf6"  
5004. 5597  
/gene="tylcvii"  
/function="biosynthesis of NDP-mycarose during tylosin  
production"  
/codon\_start=1  
/transl\_table=11  
/product="NDP-hexose 3,5- (or5-) epimerase Tylcvii"  
/protein\_id="AAD41825.1"  
/db\_xref="GI:5305795"  
/translation="MITETVRDAYRITPEIPDRHRSYLSRYETLRRATGHAIE  
IQAHVTVNRNLTGRHGTVPVPGQKIVTCVRGAARTWVLDLVGSPFTGGHVVVG  
ODAGVAVVLPDGLGIVVALADTCMYLYTREVTPGMIIDIDALDPLGLPWNLT  
EPVRSERDAAPSLAEAAAGTLPGEQCLRAPAM"

ORIGIN

Query Match 20.8%; Score 104.2; DB 1; Length 5908;  
Best Local Similarity 55.7%; Pred. No. 7.1e-05;  
Matches 199; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 142 ACGTGTACGCGGGGACGTACAGATCCACTGTCGCGCAGCGCGCGGAACTCCTGCTCC 201  
DB 2113 ACGTCGACCGCGGAGCGGGAACACAGTCGCGCGCGCGCGCGGAGAGTCCCGCTCC 2172  
QY 202 TTGCGCATGATCTGTCGCGGTGTTCCAGCGAAGACAGCGCGTGTAGTCCACCGCTCG 261  
DB 2173 CGTCGACGAACCTCCGCTGAGTATCCAGCGCAGGACGACGTCGTGTCGCGCGCATC 2232  
QY 262 GCGGTGACGCGTCGCGGTGCGACCGGATGTCGTCGCGGGGTGACCGCGCCCTGC 321  
DB 2233 GCCTTGGCGTCTCTCTCGAGACGATCGGATGTCGCGGAGTGAACCGCGCCGCT 2292  
QY 322 TTGCGCGCGGTGTCGTCGACACACCGACGAGACGAGTCCGACCGATGCGCGCAGAGTTC 381  
DB 2293 TTCTCGGGGTCACCTCCCGATGACAGGAGTCCGTCGCGATCCCGCAGTACTGG 2352  
QY 382 GTCACGTTGCGGCTCTTCCGCGTCCGCGGTACGACACCGCTGTCCTCGGCTTG 441  
DB 2353 AGGATGACATTGCTTCTGAGGCGCGCTAGCGAGAGTACGCGCGCGCGCGCGG 2412  
QY 442 AGCGAGTTACGAGCGGCGAGGTCGCTGCGGATGCCCTCGAGTCCGCGCGGAGAC 498  
DB 2413 GACCGTCCAGGAAGTCGCGAGGCGTCCGCTGCTGACCGCGCGGCGGAC 2469

RESULT 25  
AE008206/c  
LOCUS  
DEFINITION  
187 of the complete sequence.  
ACCESSION  
AE008206  
VERSION  
AE008206.1  
KEYWORDS  
GI:15158374  
ORGANISM  
Agrobacterium tumefaciens str. C58 (Cereon)  
Agrobacterium tumefaciens str. C58 (Cereon)  
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
1 (bases 1 to 12862)  
Hinkie, G., Slater, S.C. and Goodner, B.  
Complete Genome Sequence of Agrobacterium tumefaciens C58  
(Rhizobium radiobacter C58), the Causative Agent of Crown Gall  
Disease in Plants

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

Unpublished  
2 (bases 1 to 12862)  
Hinkie, G., Slater, S.C. and Goodner, B.  
Direct Submission  
Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney  
Street, Cambridge, MA 02139, USA  
Approximately 800 bp of telomeric sequence missing from the left  
end of the chromosome and 200 bp missing from the right end.  
Location/Qualifiers  
1. 12862  
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"  
/mol\_type="genomic DNA"  
/strain="C58"  
/db\_xref="taxon:181661"  
174. 1466  
/gene="AGR\_L\_168"  
174. 1466  
/note="AGR\_L\_168"  
/note="hypothetical protein slr1610 - Synechocystis sp.  
(strain PCC 6803)"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_L\_168p"  
/protein\_id="AAK88650.1"  
/db\_xref="GI:15158375"  
/translation="WNEAVPSARLSSTQHVGLCRLCNSPLVHTFVLDGMSPPCESFV  
ADAANDVEPPYPLHAFYCDCEFLVQVAPENIPEYAFSSDSWVAHAKRYC  
DVIERFALGQSFVVELASNDGILLQHLFASNPMLGIEPAAVAVKVAIGKIEPTLT  
EFFNEALATDMAARQKADLIIGNVLAQVDINDFVAGMKTLKPGCVITIEFPHIE  
KLIENQDPTTYHEHFSYSLTITKMAARHGLKVFVBEIPTHGGSURVPSHEDIR  
FPREARVDSLLARELNAGLKIEITYTFAEAVRQTKRLLSFLRLKEMRKSICAYGA  
PCKTTLNLYCGIGTDFDAVDNRNPKHGLTPGMHPIRPFVSEIHPMKDYYLLLP  
WLNKNEIQAOKMDIRNMGKFIPIPIIDISIDPKELVQ"  
1463. 2275  
/gene="AGR\_L\_170"  
1463. 2275  
/gene="AGR\_L\_170"  
/note="(AL121855) putative transferase"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_L\_170p"  
/protein\_id="AAK88651.1"  
/db\_xref="GI:15158376"  
/translation="MKVYLCGGRTIRREYSESVKPLIPLGSPINRHWVQYARY  
GHEDFICLGYKAVHKEFFLNSRPESYADCVISPGRVLELLELPDRVSLIDTGI  
WNIIGRLWAVRHKLKGMFLANTSDGLTDVDDMLAQFESKDLACFLAVRPLT  
YHLEIDSGKVRFRSSETSDMWINGVFLMRPIFDYMRGEBLVEPFRSLISED  
KLWAYKHGFWRSMDTLRDWQTLFAMVERGDMPIWGDDASMARKSSNAKLIP"  
2215. 2925  
/gene="AGR\_L\_171"  
2215. 2925  
/note="(AL121855) hypothetical protein SCP62.20"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_L\_171p"  
/protein\_id="AAK88652.1"  
/db\_xref="GI:15158377"  
/translation="MDWRECVNGTQIVRQAAAMKNMSLASAGDRSLISLCAHSDDI  
EIGVGGTVALSLKGYLLDVTWCVLSAGNERALEARSAAFLAKAASYSIELASFQD  
SYVPSOSRIKSWLLTRDKVSPDIIFTHSREDADHQRHREINRLTSNIFNHLIYEY  
IPKWDGDFGRNAYVPLSEDAAMMKVELLQHQFGTQRSKMDPDAFTFGLARLGMEC  
RAPTRPAEAPTVRKLTPA"  
2865. 3722  
/gene="AGR\_L\_173"  
2865. 3722  
/note="(AF076290) putative response regulator"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_L\_173p"  
/protein\_id="AAK88653.1"  
/db\_xref="GI:15158378"

```
/translation="MPCADTFCSGLYRQKAHRLKQAOALRRYEDIVFKPKETKNTT
GGIFVMIRSELNPEGAHPHPOAASAAAGYIGIDREMLRECLAHSLNVOYADRCI
LPSNLGEMQOARAEIGMAANLVSAWPGPSAETHRELAEIVAAVSPAAYVVLSDQD
INTILITIALGVNGYISSVSVDVCQALSLAIGLAFIPASSVMNKLFLGLTAHPA
IPMKFTSQSAIAEALRCGKANKDIALENLCESTIKVHIIRNKKGLGATNRTEVAC
KIDMAYER"
gene
3777..5939
/genes="AGR_L_175"
CDS
3777..5939
/genes="AGR_L_175"
/notes="(AF076290) Bmel2"
/codon_start=1
/transl_table=11
/product="AGR_L_175p"
/db_xref="GI:15158379"
/translation="MQSSSSGCEQGISRPISILERRERRAVIFCLASFAAATGYA
RTPIRYKAENTALDVRKYLQALPTESVSPLEPQESVPLTELDIDIGSRMAERVLIV
QKDLLEGQNTDRATNDYDQIDQITATRSVAGLGRVLSRLTALLESERQASQFRE
DPALAAQANAFGEAYIDYQIDQITATRSVAGLGRVLSRLTALLESERQASQFRE
KSLGINTGLQSLQSLSSINVELTTLQAKLAGRARLATALQKSGDGLGITVLS
SPITQILRAQARVIRIAQIDSGAVMSPERPOLSSQSLNKQIDSEIAQIIAALR
NEVEVEKOROEVEASLNDVORAISTTDMAMVQSGQLDREAAANSIYETIARYKOT
IEQDQATAEARLISAMPSPRKPTSPDAGLGLVALLGCSGLATAFILNPTDSVG
SITSLOQKTGLAVIGIRIPRISFGEPONKPK3SSGAADFYSAFDLQHLQLAESKAK
VIAPTSAGEGTFVIANLARSLAMSGVKLLLDIDANLRNPALGAEEFDVGSTSHVAQA
VSHVTLDDIIQDQASGVDDIIIAEKSDMPNVLGSKRFAGLIVEARQRYELILIEA
PSYDHLDLIRIATLSDAVSFVVRQDADGTGMKNVAKIRLVAGRPITGIILNGIRRP
RKQPPVWPRIFRDMWPHFKLRATSLRPPAVVHNER"
5943..7202
/genes="AGR_L_177"
5943..7202
/genes="AGR_L_177"
/notes="(AF076290) Bmel1"
/codon_start=1
/transl_table=11
/product="AGR_L_177p"
/db_xref="GI:15158380"
/translation="MFLTRSGSSLMIASTTFLSLAFVQIPAAADGTYKYVATGDVL
TISVDPGDTGLFPFVADGTIGYPLLGNNVVDQTVNEIGERISRLASHVANRSVA
YAVEKAPIFVGDVQKPGYEPGNVILELFPALGGREPNADQTSIGILIAAQQ
EYEDMSQLLSQDIRVRVLAELNDTEFEVGEIGLVRDRAALEKIVASETSLRFLR
LSAQDEKTLNVOQNFIQEIIDLQKSNVMOGQLELLDMVDNASKELATVTRGAASES
ALRERKEILLNVOQVLESTFLARAQONKSEVEREILELKSKEHNDATRETSLD
IMRLKKLAFSLQSMIEIGSTARVSSLEDMIQTKFFIVRQVAGTYRETADEHTQVR
AGDVVRVLSVGRDPVSAAVVTGSVN"
7426..7722
/genes="AGR_L_178"
7426..7722
/genes="AGR_L_178"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_178p"
/db_xref="GI:15158381"
/translation="MSAIVPVGVHVRORAVIILNGERGLQPVHVTDALNDIOSPP
QCDARLAQFVDVFAITIAASKIEKGFAFDGQIWIATDVQVWNLAAENGPAK"
8232..8411
/genes="AGR_L_180"
8232..8411
/genes="AGR_L_180"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_180p"
/db_xref="GI:15158382"
/translation="MPGFQSEWTPVQSGVGLAALEBDLTAPPLGPPFPPIAKPLHVEPT
FTQLFCGCVTKPAL"
complement(8789..10792)
/genes="AGR_L_183"
gene
```

```
complement(8789..10792)
/genes="AGR_L_183"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_183p"
/db_xref="GI:15158383"
/translation="MVATPEVAALGVYTTSGKMRSPNSLISILLAAALAFASGLSFA
GKPAAGVQVQSGSDPRQFPPIPEPRPLPDDADAPNPMEVGGPAIAISRT
AGGEVVMTGVDUDGRTGFEVFSQFPGRAAGGVVPLIAPITADATATATLILLPATLPSW
STYLIRPLRGDLGRSPAINRTESWIGPDAAAPGAIVSYGRNLSFNGTSQSYLYV
KPAAGWYKAIASVNPFFKVDYFVLAIPPARYEVVWHTHGGFGSGPLTDLVSQS
PWAQDKKRIIDARQFAGKAGVTDTTAAILRAIDAAATAPATLYFAGIYLVTAPLY
APANWAFEGEKNKTEIRLNKRIDQSMVVDGDDTRFENITLANRNRIAGKPLIALAS
TKNRLDSVRVDAWPAFDAQDTERLFVASELIEGSPFVAGASROVFFSDNRFSMTG
XGESVLDVSGRDPFAMTGNELVNADESDDHGGIGRFFVQAQHFSGMKNLYFEKNISR
NAASHDCTKVCNKGEOICFEIVGSKLNDPFRASNSVTFKSLDGLDADTPGDFDLV
IVGGRGAQHRHIASTFVLEQAWNVDPDTSRFRALAAATRAVIYDNTLDGRES
YAEHDSSTGVLLYGVYDVVDNSNRISMRHGMVTVALDSTR"
complement(10735..11730)
/genes="AGR_L_185"
complement(10735..11730)
/genes="AGR_L_185"
/notes="HYPOTHETICAL 34.7 KDA PROTEIN Y4AF"
/codon_start=1
/transl_table=11
/product="AGR_L_185p"
/db_xref="GI:15158384"
/translation="WSSSPSKMPPSLNGKRVVAGTGMVGSALVRLRLENCEIL
KYSRREIDLTFQYETHEQWMAAARQVIFVAAKVGGAANNAAYPADFLYTLISMNI
MKSADIGVEKLVLMWSSCTIFPKPAQFPIETENALLTGPLETNEAYALAKIALKLSQ
FYSIOYGLNCVNMPTNIYGLNDNFDQSSHVIIPAMIRRMHEAKISONKILVMTGTS
PLREFLVDDIADACFLPMKSSAHFPLINTSGREISIRNLHLIAGIYVGGQIVFD
TSKPDGAPRKLDCSRLNALGNSTVELRYGIQDLYEWWRHPKSLQSDMLRQAAAR"
complement(11714..12721)
/genes="AGR_L_187"
complement(11714..12721)
Query Match 19.6%; Score 98.2; DB 1; Length 12862;
Best Local Similarity 54.6%; Pred. No. 0.00034;
Matches 196; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 140 TCACGTGTACCGGGGGACGTACAGGATCCACTGTCCGCCAGCCGCGGGAACCTCTGCT 199
DB 1440 TGATGGATATGTCCGGAATTGGAACGATGAATTTCCACCACCCAGTTCTTGATGCTTCA 1381
QY 200 CTTTCGCCATGATCTCGTCGGCGTGTTCACAGGCGAAGAGACGCGGTAGTCCACCGCGT 259
DB 1380 TTTGCGCAACGATCTCATCTTCAGTTCCAGGGGAGAAATGAGAAACGTAGTCCGCTTCA 1321
QY 260 CGGGCGTGAAACGTCGCGGGGTGCGCACCGGGATGTGCGTCCCGGGGTGAGCGGCGCT 319
DB 1320 TGTGGGGATTTCGCTGACGCGGCGGATCGGGATATGCAATTCGCGCGGTGAGACGACCAT 1261
QY 320 GCTTCGCGCGGTCTGTCGCGACACCCAGGAGACAGGTCGCGGACCGATGCCGAGAGAGT 379
DB 1260 GTTTGTAAGGTTGCGATCGACTCGGAATCGATGAATCGTATATGCGCAATAGT 1201
QY 380 TCGTCAACGTCGCTCTTTCGCGTCGCGCGGTAGCGCACCCACCGCTTGCCTCTCGGCT 439
DB 1200 TCAGCAGCGTGTGCTCTTCCCGGTGCGCATATGAGGCGAGATCGATTTGCGCATCTCT 1141
QY 440 TGAGCGAGTTTCAAGAGGCGGACGAGTCGTCGCGGATGCCCTCGACGTCGCGCGGCGAAC 498
DB 1140 TGAGCGGATGAGGAGACAGGAGTTTCGCTTGGTCTGGCGACCGCTTCGCGGAC 1082
RESULT 26
AE009408 12952 bp DNA linear BCT 20-DEC-2001
LOCUS Agrobacterium tumefaciens str. C58 linear chromosome, section 178
DEFINITION
```

[illegible]

PDIIPTHSHDAHDHREINTSIFRNHLILEYETPKMDGDFGRNNAVPLPSDAM  
NMKVELLQHFCTORSKDMFDADTFRGLARLRGMECRAPTEFAFTVRKLTPA"  
complement(7499. .8311)  
/gene="rfbf"  
/note="synonym: Atu4797"  
complement(7499. .8311)  
/gene="rfbf"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="glucose-1-phosphate cytidyl transferase"  
/protein\_id="AAL45591.1"  
/db\_xref="GI:17743309"  
/translation="MKVILFCGGRTRIRREYSVPKPLPLGSPQIMRWVQYARY  
GHEDFILGLYKAHYKEFFNLNRPESVADCVISPGRVLELPRDVRSLIDTGI  
WNIIGERLWVPHLKGKEMFLANYSDGLTDVLDLDMIAOFKSKLACFLAVRPPLT  
YHLAIDSGKVRPSETSDMTMNGYFLMRPEIDYMRGELVLEPRLISED  
KLMAKYGSGFRSMQTDJRWQLEAVRGMVWIGDDASNAKSLP"  
complement(8308. .9600)  
/gene="Atu4798"  
complement(8308. .9600)  
/gene="Atu4798"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAL45592.1"  
/db\_xref="GI:17743310"  
/translation="WNEAVPSRLSTQHYGLCLNSPLVHTFVDLGMSPCESFV  
ADAANDVEFPFLHFAVDECFVLQVEYAPENIFTEYAFSFSVSHVAKRYC  
DNVIERFALGQSFVVELASNDGLYLHFLTNSIPMLGEPFVAVKVAIKGIPTLT  
EFNFALATDMAAGOKADLIIGNVLAQVPDINDVAGMKTLKPEGVITLEPHIE  
KLIAEQDTIYHESYFSLTIEMARRHGLKVFDEETPTGGSRVFSEHDIR  
PFEAREVSLRLAGLAKIETYPFAEAVRQTKRNLSFLIRLKMRSKICAYGA  
PGKGNLNYCGIGTDFIDFADVRNPYKHRLTPGMHPIRPVSEIPHMKDYPVLLP  
WLNKNEIQAOKDIRWCKGKFTVPIDISIIDPKELVQ"  
9905. .10432  
/gene="rfbc"  
/note="synonym: Atu4799"  
9905. .10432  
/gene="rfbc"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer"  
/codon\_start=1  
/transl\_table=11  
/product="dUDP-4-dehydroxamnose 3,5-epimerase"  
/protein\_id="AAL45593.1"  
/db\_xref="GI:17743311"  
/translation="MLFEKTKIEGLLVNTIPVDRSGSFARTFCKREFESNALVSDF  
SQHLSLSLKHKTIRGLHPQHPQETKLVSCVKGSIWDVAVDLARGSPFTYLQWGTV

Query Match 19.6%; Score 98.2; DB 1; Length 12952;  
Best Local Similarity 54.6%; Pred. No. 0.00034;  
Matches 196; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 140 TCACGTGTACGGGGGACGACAGGATCCACGTCCGCGCAGCCCGCGGAACTCTGCT 199  
Db 8334 TGATGATATGTCGGGATTTGGACGATGATTTCCACCCAGTTCTGATCTTTCA 8393  
Qy 200 CTTTCGCCATGATCTGTCGGCGTGTTTCAGCGGAAGACGCGGTGATCCACCGCT 259  
Db 8394 TTTTCGCCACGATCTATTCTTCAGGTTCCAGGGGAGATGAGACGTAGTCGGGTTCA 8453  
Qy 260 CGGCGGTGAACGGTTCGGGGTTCGACCGGGATGTCGTCGGGGGTGAGCCGCCCT 319  
Db 8454 TGTGCGGATTTGCTGACGGCGGATTCGGGATATGATTCCTCGCGGTGAGACGACAT 8513  
Qy 320 GCTTGGCCCGGTGCTGTCGACACCCAGAGACCGAGTCCGACCGGATGCGCAGAGT 379  
Db 8514 GTTTGTAAAGGTTTCGATCGACTCGCAATCGATGAATCGGTACCTATCGTCCGCAATAGT 8573

Qy 380 TCCTCAGGTGGCGCTCTTTCGCGGTGCGCCGCTAGCCACACCGCTTGCCTCGGCT 439  
Db 8574 TCAGCAGGTGTTGCTTTCGCGGTGCGCCGCTAGCGCAGATCGATTTGGCATCTCT 8633  
Qy 440 TGAGCGAGTTCAGCAGCGGAGCAGGTGCGTGGGATGCGCTCGAGTGGCGGCGAAC 498  
Db 8634 TGAGCGGATGAGGAGGACAGGAGTTTCGCTTGGTTCGCGCACGCTTCGCGAAC 8692

RESULT 27  
SCO939105/c  
LOCUS SCO939105 291000 bp DNA linear BCT 11-FEB-2003  
DEFINITION Streptomyces coelicolor A3(2) complete genome, segment 2/29.  
ACCESSION AL939105 AL109974 AL117385 AL117387 AL121596 AL121600  
VERSION AL121719 AL121849 AL121855 AL122662 AL132707 AL645882  
KEYWORDS AL939105.1 GI:24418991  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1  
REFERENCE Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,  
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,  
Huang, C.H., Kieser, F., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,  
Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,  
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,  
Taylor, K., Waters, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,  
Parkhill, J. and Hopwood, D.A.  
Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)  
Nature 417 (6885), 141-147 (2002)  
21996410  
PUBMED 12009953  
REFERENCE 2 (bases 1 to 291000)  
AUTHORS Bentley, S.D.  
TITLE Direct Submission  
SUBMITTED (09-MAY-2002) Submitted on behalf of the Streptomyces  
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
On or before Oct 29, 2002 this sequence version replaced  
GI:20520891, GI:20520894, GI:20520895, GI:20520772, GI:20520896,  
GI:20520897, GI:20520883, GI:20520899, GI:20520902.  
FEATURES  
Location/Qualifiers  
1. 291000  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="A3(2)"  
/db\_xref="taxon:100226"  
213. .3092  
/gene="SCO0305"  
/note="synonym: SC5G9.14"  
213. .3092  
/gene="SCO0305"  
/note="SC5G9.14, hypothetical protein, len: 959 aa;  
unknown function, similar to TR:O54182 (EMBL:AL021411)  
Streptomyces coelicolor hypothetical protein (913 aa),  
fasta scores; opt: 698 z-score: 776.6 E(1): 0, 27.0%  
identity in 972 aa overlap. The C-terminus shows very weak  
similarity to eukaryotic beta-transducins e.g.  
SW:GBB4MOUSE (EMBL:S86124), gnb4, Mus musculus guanine  
nucleotide-binding protein beta subunit 4 (340 aa) (23.0%  
identity in 291 aa overlap). Contains possible hydrophobic  
membrane spanning region"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="CA355662.1"  
/db\_xref="GI:5881863"  
/db\_xref="GOA:Q9RL12"  
/db\_xref="SPTREMBL:Q9RL12"



```

/translation="MSDAVEAEAEARAVPVGAKORDAFISYQSQKNAPLAKQLQRLGE
RLAVSRPWRMSVFDVTSPLPNDHMQSIORELARSRYFILLASPEAAASPWVARE
VYWLKRETDHLLIYASGEIVWDRAGDFWRKTSPLRNLGHTAAPLWVLDRE
IGERSHSMRYPPRAAVATLAPLHGEPEKELEHNDPQLRFVRLSGVALLVAL
LVLAVYGFDSQSRQAALQSARTSASQAARSQQLGTSNQAQYALVAEETST
PSRRLAQAATAAPFARRRLRADDSVTDMSQASPAATDVVLSDGTTAAYRSVFD
DRAVRLYDVSARQSGVVRABGTFRALSRDGRVLRAMEVILNRVQMDTRSGRLLSGT
ATGHTENLPHVGHGLHALASPDGRWAAASHYTPGEAFVAVWRASDGREITRVPSVD
ARVGLGFSADGSLTIVDQDRREVRFRDTSARWEAARSLPGMPGSTAEALRYSEVLL
FDGAKALTRTSRAEAWDLEQRRLIARSLGPEYEDMTVADDKGIPLGGQGTIVRLY
DALRPGTGLRLTRPASTIASADGTQVVVASYGELNLTFTTVRRGQOAVDEKDFV
TFGLDLPDGMVAHREGTEHEDPATGRGLGVIYANLSVLEDTAVTLSDGKRYOG
MQSVGEFDEGRDWDLTGRIDTCCTAVGSSSPRRPGLSILPGDRIYVYGEWG
AVELLTRICTVSTEDSGRFALSGDRRLTILVLDNDGVDTWKWDGDHAFAPAAKITL
PSAHPVDMVSVDHEGQAAPFSRSHVYVVRDLSGRVEATGLYLPDKTDVAFPSRDGR
LLPQGFMTSTGRVRLDAGSGDQLDITWTDPTGTSVQNEGMVQVPGPADDILTLP
DHKVVRTVGVGAHDLICLVSQLPARSDRYLKDLDVEDPCR"
/complement(3056..3961)
/gene="SC00306"
/notes="synonym: SC5G9.15c"
/pseudo
/complement(3056..3961)
/gene="SC00306"
/notes="SC5G9.15c, possible pseudogene, len: 302 aa;
similar to hypothetical proteins e.g. SW:YCYJ_ECOLI1
(EMBL:AE00230), YCJY, Escherichia coli hypothetical
protein (310 aa). GC frameplot and codon usage plots
indicate a coding region which lacks a start codon"
/pseudo
/codon_start=1
/transl_table=11
/product="putative pseudogene"
4084..4971
/gene="SC00307"
/notes="SC5G9.16"
/notes="synonym: SC5G9.16"
4084..4971
/gene="SC00307"
/notes="SC5G9.16, possible DNA-binding protein, len: 295
aa; probable CDS suggested by GC frameplot and positional
base preference. Similar to possible DNA-binding proteins
from S. coelicolor e.g. SCJ9A.15c (cosmid SCJ9A), S.
coelicolor possible DNA-binding protein (290 aa) (39.3%
identity in 298 aa overlap). Contains probable
helix-turn-helix motif at residues 35-56 (Score 1197,
+3.26 SD)"
/codon_start=1
/transl_table=11
/product="putative DNA-binding protein"
/protein_id="CAB55664.1"
/db_xref="GI:5881864"
/db_xref="GOA:Q9RL11"
/db_xref="SPTREMBL:Q9RL11"
/translation="MSSELGDFLKARRELSPPTAGLPADGRVRVPGRLREEVAVLAS
TSPGHVTLBOGRRRASPVLDALARLISADERAYLVFLSGDKAGRPGRQVR
PQLRLLEDLHTFALVLRCTDILANPPAALFTDFALLSRDPRFVLRVPRDPAV
RALYTDWFMARCAVQRMEARDPRLSALVALSIAQDPDPTFWKDHVAVRG
AGTKDLRHPVIGELFLDWSALTDTADPQQLIALTAAPGTPSHGLRALRDLGAVTSG
RTGPGVGGGGIAGASQA"
5026..5429
/gene="SC00308"
/notes="synonym: SC5G9.17"
5026..5429
/gene="SC00308"
/notes="SC5G9.17, possible membrane protein, len: 467 aa;
unknown function, probable CDS suggested by GC frameplot,
positional base preference and amino acid composition.
Contains hydrophobic, possible membrane-spanning regions
near the C-terminus. Contains TTA (leucine) codon;
possible target for bida regulation"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB55665.1"

```

```

/db_xref="GI:5881865"
/db_xref="SPTREMBL:Q9RL10"
/translation="MPRSDAVRVTRLLMYHLTPQVHVDPDRDLRLPLSEALGLGRTF
ARVEAGALLPRLVLRVREITLPFNAARVDGADITLAVTPQDPLFLVDVEVPDEA
GDSPTTAETLYATWQDRSAIRVGGVPLTAWLDERLAYGLTDRTPPLFAGRNVHQ
SVFGAAGRELLAAHPVGDVGVVITVRLTSDGTGATLGRPETINNAQTQM
VANGRGSLTVGVAPPVPAAGVAAAGITVRAVAHRIRROQAFAALDELDPAAAVETTD
AARALVAQLSHRLNDLQDLSESVYVADTILVPELLVESHASRLPSLSALANT
SRIVERNVIAITRRVGLAQAQYADDRVLAVIAAGSLALPALLAVLAVGUNS
SDVPSISILDFQHYGAYIAWSPFLLILVGLLARISRLPAPDAATSERTRER
FRSLRSWLGRSRPPCR"
5434..5436
/gene="SC00308"
/notes="TTA (leucine) codon; possible target for bida
regulation"
/complement(6537..7625)
/gene="SC00309"
/notes="synonym: SC5G9.18c"
/complement(6537..7625)
/gene="SC00309"
/notes="SC5G9.18c, probable amide hydrolase, len: 362 aa;
similar to 6-aminohexanoate-dimer hydrolases which degrade
N-(6-aminohexanoyl)-6-aminohexanoate e.g. NP:P75007
(EMBL:X02336), NylB, Flavobacterium sp. (strain K172)
6-aminohexanoic acid linear oligomer hydrolase (Eri') (392
aa), fasta scores; opt: 531 z-score: 598.0 E(): 6.1e-26,
34.1% identity in 323 aa overlap"
/codon_start=1
/transl_table=11
/product="putative amide hydrolase"
/protein_id="CAB55666.1"
/db_xref="GI:5881866"
/db_xref="GOA:Q9RL09"
/db_xref="SPTREMBL:Q9RL09"
/translation="WPSRRVWRGCGDLPASPFGALDGLVGTSGVTRITLQDLA
EARTDALVLRGLVHEEYRHYEAHVPHFNASAKSYLGLLAATLAHEGLRDRGAR
VARYPELSTGAFGDAQVDHLMEMGTMSYAGRPYDKALEAQRFAVLALRFPYGS
GPATIREHLATARGEGTEFRYENGVEALSLVLRITGLTSALMSEMIWSRIGA
BEDAYVLDGAEACAGGFSATARDVARIGELRCGGAQVDRQVLPAVTPARSVP
DCPRRVFFPAAPASDPATLSYDLWLMDLWITLNDPYGFSMASGHGRLFFVSPALDLVVVH
YASQVSPAVPQVPLVQAPLSIGAHANA"
/complement(7816..8466)
/gene="SC00310"
/notes="synonym: SC5G9.19c"
/complement(7816..8466)
/gene="SC00310"
/notes="SC5G9.19c, possible tetr-family transcriptional
regulator, len: 141 aa; shows similarity to the N-terminal
half of TR:Q24741 (EMBL:AB001683), FarA, Streptomyces sp.
strain PRI-5 receptor for an extracellular regulatory
factor (IM-2) (221 aa), fasta scores; opt: 154 z-score:
193.4 E(): 0.0021, 36.7% identity in 79 aa overlap and to
putative transcriptional regulators. Contains Pfam match
to entry PF00440 tetr, Bacterial regulatory proteins, tetr
family"
/codon_start=1
/transl_table=11
/product="putative tetr-family transcriptional regulator"
/protein_id="CAB55667.1"
/db_xref="GI:5881867"
/db_xref="GOA:Q9RL08"
/db_xref="SPTREMBL:Q9RL08"
/translation="MDMKPDRRVRRTRAAALRKALVELVDKQFQAVTVETTERADI
GRATFYVHRVDKEDLLVGLVRDLAQDRERLPAVRQAHAQGTGLPFIFFHAQOER
GVYQVILREGDGRALRETFELIRSHAEMAFDRARQLGVSPRIPLGIVARATGELI"

```

Query Match 19.4%; Score 97.2; DB 1; Length 291000;  
Best Local Similarity 54.5%; Pred. No. 0.00019;  
Matches 195; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 141 CACGCTAGTCGGGTATGGGAGACCGAGCGCGCCGCCACTCGTGCAGGAGGAG 115405  
Db 115464 CACGCTAGTCGGGTATGGGAGACCGAGCGCGCCGCCACTCGTGCAGGAGGAG 115405

```
QY 201 CTTCCGCCATGATCTCGTCCGCGTGGTTCCAGCGGAGAGCAGCGCGTAGTCCACCGCGTC 260
|||
Db 115404 CTGCTCGACCAAGTCTCTCCGCGAGGTTCCAGCGCAGAGACGATGTCGCGCGGTC 115345
|||
QY 261 GGCCTGTGAAGCGTCCGCGGTCGCGACCGGAGTGCCTGCGCGGGGTTGAGCGCGCCGTG 320
|||
Db 115344 GCGCGCGATCCGCTCCGCGCGACAGGATCGGAATCGCGGTGCCCGGTGAACCTGCCGTG 115285
|||
QY 321 CTTGGCCGGGCTGCTGTCGACACCCAGGAGACAGTCCGACCGATCCGACAGGTT 380
|||
Db 115284 CTTGTAGGATTCGCTGCGACCTGTACGCGACAGTCCGCGCGACCGCGACGATGGTT 115225
|||
QY 381 CGTACGCTGCGCTCTTCCGCTCGCGCGTACCGACACCGCTTCCCTCGCGCTT 440
|||
Db 115224 GAGCAGCGTCTTCCCTTCCGCGCGCGTACCGACACCGCTTCCGCGCGCTGCGC 115165
|||
QY 441 GAGCGAGTTCAGCAGCGCGAGGTCGCTGCGGATCCCTGACCTCGCGCGGAC 498
|||
Db 115164 CGCTCGATCAGAACCGCAGCAGGTCCTCCCGCACCTTGGCCACCGCGGACGAGAAC 115107
|||

RESULT 28
LOCUS AY048670
DEFINITION Streptomyces globisporus enediyne antitumor antibiotic C-1027
ACCESSION AY048670
VERSION 1
KEYWORDS 85163 bp DNA linear BCT 05-NOV-2002
SOURCE Streptomyces globisporus
ORGANISM Streptomyces globisporus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomycetes.
1. (bases 1 to 85163)
Liu, W., Christenson, S.D., Standage, S. and Shen, B.
TITLE Biosynthesis of the enediyne antitumor antibiotic C-1027
JOURNAL Science 297 (5584), 1170-1173 (2002)
MEDLINE 2171413
PUBMED 12183628
REFERENCE 2 (bases 1 to 85163)
AUTHORS Shen, B., Christenson, S.D., Liu, W. and Standage, S.
DIRECT SUBMISSION
TITLE Submitted (27-JUL-2001) Chemistry, University of California, Davis,
One Shields Ave, Davis, CA 95616, USA
FEATURES
Location/Qualifiers
1..85163
/organism="Streptomyces globisporus"
/mol_type="genomic DNA"
/strain="C-1027"
/db_xref="taxon:1908"
complement(8..658)
/notes="ORF(-7)"
/codon_start=1
/transl_table=11
/product="hydroxylase-like protein"
/protein_id="AA06648.1"
/db_xref="GI:24575077"
/translations="MGMTGARPLVVGAGLACTATAILLFARPLVLLERRAA
YRAGVAYHRDGNPDHVFNIQAGRSVPREDVLDFFINWAEADRRDPRWASWKF
TEQGPAPRPLFODYLDARLVARESCGVLVEADGEALDARSHDRCEFTVGLTP
YLEGLRPGPLPOTQILDADHVVLATGLELKEPPFAAGVAGHPFVRNPSYAFGL"
complement(930..1478)
/notes="ORF(-6)"
/codon_start=1
/transl_table=11
/product="viral infectivity potentiator-like protein"
/protein_id="AA06649.1"
/db_xref="GI:24575078"
/translations="MKRIPVHTTIDTAPAGGDTILRLKKRFRGLVNIHGGMAHSPVL
ETVAITGVAHEGTFDARTREAILAVGADVACQAAHTVSXVAGFTLEETVAI
RRGTGDDVKLEALVQVAREIAGEVGEASDASWNAVAQGWTDITELAEVFEVAVNLY
TNYFNHVARTEIDVPGVPDIDS"

complement(1649..2713)
/notes="Orf(-5)"
/codon_start=1
/transl_table=11
/product="truncated methionine synthase-like protein"
/protein_id="AA06650.1"
/db_xref="GI:24575079"
/translations="MTIATEPIGSLPRSATLLHALAAHQGTLDTATLAKQOEQAVAD
TLRLERVGSPVLVDGQSQSFATYPTVGLDTPDGAVIPADGHTROLPSITQSP
FRYOVRAETVLRAARQLTDRPLKQAVIAPSAISLLYPATIEGVPRQFRLDLADEAE
ADIRGCDAGAHVQLDFTGRSLKLDPSGVLDDDFIALNNVGLGRFSAEETARIGV
HTCGQDQSTHSLIDYABLLPKLFQKAGNFYLEAGADPERVLSIVRDLHPAA
RVFLGVTDPIPVVETPEQVRDRILLAAIPIVQLGTCDDCGSPFADDTSTTRDLA
PAKTEARVGTALAEVIGL"
complement(2851..3237)
/notes="ORF(-4); similar to SCP1"
/codon_start=1
/transl_table=11
/product="transcription factor"
/protein_id="AA06651.1"
/db_xref="GI:24575080"
/translations="MSSLTDSLIVHEAKVWVLEDPGLDYVRQALDKTPRRKNKPRY
ARDGMIGYIELGADAEADPSGLYRRRVFLLPHDRSDSPEGVYRGAFGEAVDPRT
IEPNRVGEKTPRSQLTGSTVAATGS"
complement(3442..4971)
/notes="ORF(-3); similar to SCP1"
/codon_start=1
/transl_table=11
/product="putative primase"
/protein_id="AA06652.1"
/db_xref="GI:24575081"
/translations="MSSSAGSPRDATAAQAQMLAETTPAPALLPVOAAPEAPVAA
TAWAGIIPASLTDRGNAKLFLVLYRQFHEVGLGKYSWGVKRAAGGKAAALWAA
GEMAEFPGSDPGLFTDRHLHKKRLLSTTGMKALLTQAKASPDLSLDPDITLDDP
YALCTPQVDVLRNMRKPDPTDFHSRATSAPQDIPTRWHRFLEDTPGSAEGR
EMIDFLLHLLGYSITGDVQAQVLPFHGQKNGKSVLLDVMIIQILGYDAAPGFLM
DRGAYSEHSTELTGHRLVITCSLKNDFRDEARVLLTGDKIKARRRQYFSP
TPTKLWLLGNHREVEVTCGFAWRIRLLPFRIVPDERKIDNLAVELVDEGPGIL
HMLTEGARYLATRDTLAGDPRVIATSAVANTEDHIGRFLAECLLDHPENSELVEQ
GLLYTSYSTWCNASEGIRPCTARAFATRVQEVGLASPADMIKSNCRKFYFNLAAD
E"
5982..7478
/notes="ORF(-2)"
/codon_start=1
/transl_table=11
/product="glycerol phosphate ABC transporter"
/protein_id="AA06653.1"
/db_xref="GI:24575082"
/translations="MRALPQSGGLRMAGDAPTAAGNRCPPTPIQRFLLILRQTLQAP
EAKNGSPQEQEERIVLDVNLANYPTDYDRDFLAPLRERAAEFRAHPRYRVDIN
GHDFWTIPKVARATAEGRPHIAGYATDSQLARDARRPDGKFPVTSVEAALAGRT
ILGHPVVVEDLDPVVRDSYSPGELVSLPTVTMLCYANSLLARAGVPELPRTWDE
VEAACAQAVSDGPGGHGITWANDGWVQQAVALQNGVLTQDNGRSGSATVDTVSD
EMLDVWRVTHLHSGHYLTGSPDKAGFAEFVQOKVAPTPDSSKAARELIQAGAQ
AGFEVAFPLPRNKAAPVAGQVSGDSIMLAAGLDETQGLLALTQYLISPAADW
HRTNGFVPTGAAGELLEATGFRFPQQRVAGQLKASDKSPAAUGLGLDFAVNE
VITAMDDVLRSGADPAKAFABAGVAAQQLLDAYNARNRSGSGTSPAV"
complement(7573..9927)
/notes="ORF(-1)"
/codon_start=1
/transl_table=11
/product="Jvra-like drug resistance pump"
/protein_id="AA06654.1"
/db_xref="GI:24575083"
/translations="MYGQENSVSDTDTAGAGTSGPGGIRIAGARHNLKDVSL
TIPRQITVFTGSGSGKSIPTVAEQRLNSTFSWIRNQLPKYERPOAEAE
NLTPVIVDQKPVGNARSTVGTVDQPMIRALFARFGTTKDDQPSLGSVAFSFD
PQGCPCDGLGQSIADLDKMLDRSLDDGAVLFPYKVGSPDWIOWKSGSLDPA
KPIAEYSNEELTLIRGTGKVTLTSTPOTNYEGLARFERNLKRDLALSDDR
REVERFVTDGCFSCRGARLNAAALDVRIDGKNIADYSNEVRDLTEVLAGVTEPAA
APLAKARTALERIVISIGLYLTIDRPTADLSGGEGQLKMRHLGSLAGLITIFDR
PSIGLHPRDGLRLDLRLAKGNTVLVWHDVIEIADHIVDVGFGAGVHGEIV
PEGSFALKRKAKTRTGALRRGARVKEBVRPPTGTGTENADLHNLKQSVAVPTGVL
```

```
TAVTGVAGSGKSLISGAFMEAHDPDAVFDQSAIAASSRTPVSVYGLMDPLKLFPAK
BTGNASLFSNKSCEBQCGGVIIITELAFMDPVTHCGVCEGRFPEVLEHQLR
GKSTADTLEVPADPAFFTEKALLPKRALVDVGLGYSLGQPLSLSGGEGLRIKL
ADQLRGCTVTGLDEPTGLHMSDVTLLKLLDGLVEAGNTVVVIBHNLDDVVQADWI
IDIGDGRGEGEIVFTGTPKDLAADSLSLTGYLRRLHKAQAG"
complement (9982. .11349)
/note="ORF0"
/codon_start=1
/transl_table=1
/product="Na/H efflux pump"
/protein_id="AAL06655.1"
/db_xref="GI:24575084"
/translation="MAVLAATSPIGGRLTVFLQVGLLLACAYGLRLGTRIGLPLP
VGLTAGVILGPTLLGOLSPGLSGRLPREDISOHLDDAFQCGVLLVLAAGADPP
RILKRGGLAARVSLAGLLPLGLGIAHGLVPSALLADSGERGVRALPLGVCMVTA
LPVIAKLADNLTHRVNQQLIAHAFDDARWMLLALVTLASGAGGFVITITNA
WTVFVAACVAGPPIGRUSRTGDSRPVPSAVTVGVAVVLYGALTAAAGMALGVL
FVAGATLHIAPLVRLPRTLVMAPFVPLGSLRMDLTALAESEVLLTGLVLL
VATKPKAGAAVAARSQMSRYEGLAGAGNSRGMIEVIALVGLRIGVLDVTFTI
IVLVALITSVATPILRWASSRIVLEEDETERDRLAGMTEPALSGCPAPKAREEK
TPDTS"
complement (11351. .12835)
/note="ORF1"
/codon_start=1
/transl_table=1
/product="alkylhalidase"
/protein_id="AAL06656.1"
/db_xref="GI:24575085"
/translation="MDVSAOYDVLVGGPAGSTVSTLVKRGHVLQLEKETFPVQ
IGESLPSVIGIAHLGVDELKAAFTIKHGTPFKGANPEPWTDFVSRMRMGA
TGYAYQVERMKFDQLLDNARRHGVRENSDVLDDVLAEDGRVGRVRYRDSGEH
VGSFVFDASNTGGKLGKAGKEYSFPFNALFGYFNGKRLPNSGNIPCTFF
EHGFWYTPLSGDTSGAVVNRDSASLVQDPEKAMESFTAAEPLAEKLSAATRT
EGPYGLERVKDWSYNTKFWAPGMSALVGDAAFCVDFVSSGVHLATYSGLLAASLN
SCLDGSVETSAFEPFRARYRLEYGRFYELVGVFMDHHSDDSYFMOARKISNTADSD
LQSFVELIGVSSDEFAATARORFSEASDEMAGMTHTAAGDADGPHGAVPGKVGCR
LMEQSAALQARALLGDEAGRPVQGGLAASDTGLGVNFGARS"
13012. .14079
/note="ORF10"
/codon_start=1
/transl_table=1
/product="dNTP-glucose synthase"
/protein_id="AAL06657.1"
/db_xref="GI:24575086"
/translation="MKALVIGSGSGSTRURPISYAMPKQLVPIACKVLEVLVDNIRML
DIKEVAIVGWAQAIIEAMGDSRFLRLTYIQEQLGIAHCKVLARDPLDEDDFV
LYIGIMLDGLDLSAQADHLPAPRIIVRPDPDPAFGVIELDGEGRVLRVLEKPR
EPRSLAAGVYFFADVRAVDATSPSRGGELEITDAIQLLQGLFVEAGRYTDYM
KDTGRVEDVCEVRMLGRALQVSGEVDPESELGVAVVVEGARVTRSRVVGPAVIG
ACTVVEDSQIGPVASIGRCTVRASRLSDSIVLDDASILAVSGLHSLIGRGARIAPC
ARGEARHLRVGDHVQLEIAA"
complement (14212. .14643)
/note="Caga"
/codon_start=1
/transl_table=1
/product="C-1027 apoprotease"
/protein_id="AAL06658.1"
/db_xref="GI:24575087"
/translation="MSLGRMSRRARFGVAVASTGLAAASQVAFAAPAFSVSPAG
LSDGQSVSVSGAAGATYTAQCAPVGGQDACNPATFTTDDAAGAAFSFVVRK
SYTGSTPEGTVPVSDCATACNLGAGNSGLDLGHVALTFG"
complement (14690. .15922)
/note="ORF8"
/codon_start=1
/transl_table=1
/product="amino transferase"
/protein_id="AAL06659.1"
/db_xref="GI:24575088"
/translation="MSGGPVRLLDLPYFSEPLHVGPRNVGSRDLRMERIDGALERL
ATAALDWTGAVPVCLEDEGTADVAHVERLIGRTRAILDWHVGRPARIDELTK
LAAHGLHLIPAAHAPGCTYRSKPIGFGTAELISFOATKFNVSFGAIVTDDAL
ADRLARVHQGLNAHEHTGSGTVAMHEISAAGLTLSLESDHTFATNRNRYLEYQ
```

CDS

CDS

CDS

CDS

```
YLDLGPVVRVPEQPNELSNQYVYIEVDVAVRAGLHRLDELOAVLQRHNVLARAYFSPG
CHCEPYASDLARHAPDPLPKVIALTERVLSLPTGTAVGPEVGVCRILLRAADVDSA
VPEIHSTEGDAGGRPAR"
complement (15919. .16653)
/note="ORF7"
```

```
Query Match 19.4%; Score 97; DB 1; Length 85163;
Best Local Similarity 55.1%; Pred. No. 0.00029;
Matches 190; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
QY 154 GGGAGGTACAGGATCACTGTCCGCGAGCGCGGCGGAACCTCTGCTCTTCCGCAATGTC 213
Db 16699 GGGATGGGAACCACTTCCCTCCCGCTCCAGAAAGCGCGCTCTTCCGCGAGACC 16758
QY 214 TGTTCGGCGTGTTCAGCGAGAGAGCGGTAGTCCACCGCTCGGCGCTGAACGCG 273
Db 16759 TGTTCGGTGTAGTTCAGCGAGAGAGGTAGTTCGGCTCGTGGCAGCAGCTCC 16818
QY 274 TCCGGGCTCGCACCGCGGATGTCGTCGCGGGGTGAGCGCGCTCTGCTTGGCCCGCGTC 333
Db 16819 TCCGGAGGAAGACCGCGATGCGTTCCCGCGCAGCAGTTTCCCGTCTTGAGGTGGTG 16878
QY 334 GTGTCCACACCCAGGAGACAGGTCCGCGACCGATCCGCGAGAGTTCGTACCGTGGCG 393
Db 16879 GTGTCCCGCAGACGATGATCTGATCCGTACACCGCAGCGCATCAGCACTGGTTC 16938
QY 394 CTCTTCGCGCTCGCGCTTACGCCACCCCGCTTGCCTCGGCTTGGCGAGTTCAGC 453
Db 16939 CCCTTGGACGGTCTCCGTAGCGCGCACCGGTGCGCTCGCGCGCAGACCGCGAAG 16998
QY 454 AGGCGCAGCAGGTTCGTCGGGATCCCTCGAGTCGCGCGCGCAAC 498
Db 16999 AGCGTACGATCGTTCGTCACGCGCGTCACCGCTCGCGCAAC 17043
```

```
RESULT 29
AF170880/c 25617 bp DNA linear BCT 30-MAY-2003
LOCUS Streptomyces spheroides novobiocin biosynthetic gene cluster,
DEFINITION complete sequence.
ACCESSION AF170880
VERSION AF170880.3 GI:31212900
KEYWORDS Streptomyces caeruleus
SOURCE Streptomyces caeruleus
ORGANISM Streptomyces caeruleus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 25617)
AUTHORS Steffensky,M., Muehlenweg,A., Wang,Z.X., Li,S.M. and Heide,L.
TITLE Identification of the novobiocin biosynthetic gene cluster of
Streptomyces spheroides NCIB 11891
J. Biol. Chem. 275 (28), 21754-21760 (2000)
JOURNAL Antimicrob. Agents Chemother. 44 (5), 1214-1222 (2000)
MEDLINE 20233587
PUBMED 10770754
REFERENCE 2 (bases 12457 to 14040)
AUTHORS Steffensky,M., Li,S.M. and Heide,L.
TITLE Cloning, overexpression, and purification of novobiocin acid
synthetase from Streptomyces spheroides NCIB 11891
J. Biol. Chem. 275 (28), 21754-21760 (2000)
JOURNAL 20347301
MEDLINE 10801869
PUBMED
REFERENCE 3 (bases 1 to 25617)
AUTHORS Steffensky,M., Muehlenweg,A., Wang,Z., Li,S.M. and Heide,L.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-1999) Pharmazeutische Biologie, Pharmazeutisches
Institut, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen
72076, Germany
REFERENCE 4 (bases 1 to 25617)
AUTHORS Steffensky,M., Muehlenweg,A., Wang,Z., Li,S.M. and Heide,L.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2000) Pharmazeutische Biologie, Pharmazeutisches
Institut, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen
72076, Germany
```

REMARK  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Sequence update by submitter  
5 (bases 1 to 25617)  
Steffensky M., Muehlenweg A., Wang Z., Li S.M. and Heide L.  
Direct Submission  
Submitted (30-MAY-2003) Pharmazeutische Biologie, Pharmazeutisches  
Institut, Universitaet Tuebingen, Auf der Morgenstelle 8, Tuebingen  
72076, Germany

REMARK  
COMMENT  
FEATURES

Sequence update by submitter  
On May 30, 2003 this sequence version replaced gi:11921116.  
Location/Qualifiers

source

1. .25617  
/organism="Streptomyces caeruleus"  
/mol\_type="genomic DNA"  
/strain="NCIB 11891"  
/db\_xref="taxon:195949"

gene

1. .1908

CDS

1. .1908

/gene="novA"  
/gene="novA"  
/notes="ABC transporter"  
/codon\_start=1  
/transl\_table=11  
/product="NovA"  
/protein\_id="AAF67494.1"  
/db\_xref="GI:7688707"

/translation="MKSAIATWKPDRPPDPLPBPAAQLRIIFRLFRPGRGLAVWG  
LIVGASLVAVASPFMRILPTAIPQRTGLSLALGMILTAVLSSVFGVQTLIS  
TVGQVMDLRTAVTAQRLPFLAFFTRTRIGEVQSRIANDIGMQAIVTVSTLSL  
SNLTAVTAMALDWLTIVVSLLLPVFVIGRRVGRKRKITLOROKOMATWAA  
TVTESIVSGILLGTMGADSLTSPFAESRLVDLEVRNMGWRMSVIGVWAA  
MEAVITWAGFVLSGCTVYISGTLVAFVSLQGLFRPAVSLIPGVQMTSLALFOR  
IFEYLDLPITPERPVALDKVRGEVDFDVFSTEEKDGNLHGLDLPAGSGLA  
VFGTSGKSTLSYLPRLYDVTVGGVLLDGDVDFADFTLARAAGVVSQETYLPHA  
SVADNIRFAKPATDEIEKAAQAQHEHIVTLPGDYTLVGERGYPSGGEKORLA  
IARTILRDPVLVDATSLDTRTEHVAQQAIDSLSEGRITITIAHRLSTVRDADQI  
VULDAKRIERGTHEELIDRGRYAALVRDRGAPAPVPVARDGKRRRCVRRAS  
RRSGDRBRAGS"

gene

2160. .3014

CDS

2160. .3014

/gene="novB"  
/gene="novB"  
/notes="similar to aminodesoxychorismate lyase"

/codon\_start=1  
/transl\_table=11  
/product="NovB"  
/protein\_id="AAF67495.1"  
/db\_xref="GI:7688708"

/translation="MNEPRRAASGRPLTRRGRVYLIVAAVAAVAVLPLRLT  
GGTPEGPKTIVSLVPEGRATQVYSDVKLVGPPTAKTATGARLEPCAARHN  
PEGYLPATVPVDEGATPAGLIGYVDYTAQRFSGERTEGARRNGSVYQVTFIASI  
VQAEADTPADMGRVARVHNRLADNALQMDSTLYALNRSLDITHTSPKTASPYNT  
YESKGLPPTPIGNPGEAMNAALNPTRGDWLIVTVVAPGDTFTADYAEHRNVREFN  
ANKSTPTD"

gene

complement (3030. .4088)

CDS

complement (3030. .4088)

/gene="novC"  
/gene="novC"  
/notes="putative oxidoreductase"

/codon\_start=1  
/transl\_table=11  
/product="NovC"  
/protein\_id="AAF67496.1"  
/db\_xref="GI:7688709"

/translation="MDVVVITGAGQAGLSSAYHLRRSLVDPDGTSSPSTTPRPGGAW  
QFWDLSITVEKIHGMALPGMELTGADGSPSSSEVIGGYTAYEKEDLRVHRPVYDVT  
AVREGDGLRVETSGVWSARALINATGTWDRFPNRPYFGQETFRGRLHTADYFGP  
GEFAGURVVGAGSGTQHLMEIADVAATITWTRRPVFRREGPPGEEGGAVALV  
DERVRAGLPPRSVVSVTGLEVNEAITRARERGVLDLPMFDRIITPSGVVDDGRHVDA  
DVLWATGFPFPAIDHLAPLRCASPAAGSGWRTRVPYATSATSWSGNSASTIGANRAG  
RAAVRETRALLERDPLV"

gene

4196. .4627

CDS

4196. .4627

gene

/gene="novD"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="NovD"  
/protein\_id="AAF67497.1"  
/db\_xref="GI:7688710"

CDS

/translation="MDLMSADFTTIVSLTTGTDEVTDLTRVCEQFLHESAAGRDGL  
LNFVPHATAGVAVIETAGSDDDLLAALRTLLPADNRWRHRHSGSPGHRDHVLPAV  
PPHATLPVIAGRLLELTWQSVCLVDVNTDNPDRYVRLSFLG"

gene

4808. .5461  
/gene="novS"  
/gene="novS"

CDS

4808. .5461

/note="similar to ImbU protein from Streptomyces  
lincolnensis"

gene

/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="NovE"  
/protein\_id="AAF67498.1"  
/db\_xref="GI:7688711"

CDS

5536. .6624  
/gene="novF"  
/gene="novF"

gene

/gene="novF"  
/note="putative oxidoreductase; similar to prephenate  
dehydrogenase"

CDS

5536. .6624  
/codon\_start=1  
/transl\_table=11  
/product="NovF"  
/protein\_id="AAF67499.1"  
/db\_xref="GI:7688712"

gene

/translation="WRTAVIIGTGWICTIGLARLKQGVSDVYLMDSPTVALRAEAVG  
ACTAEPPEPTVDLAVAVPPVPAHVAVPIASHQSRGTARFYVDVAGVKYSTRELDALAC  
DLATVVGHPVLVPGSGPLAARGDLFDGRFPAVLPAVATDAALNRLALVELACGAI  
PVLDAEAHRAIALGTLPQIALTLVAARLTADSGALRLGLGSVWSEIPLQVGVDSA  
TSWTQVLAANAAPVVGELKSLDLSLETLRGADGDSLAEPDGLLEFIQRGID  
GSRVPGRYGIPPTETALADVDSVDDPABLARLFDVAGVGVVVRGIDISQRPDSDP  
RTVTISVTPRDAENLLHELRRKWPANS"

gene

5730. .7686  
/gene="novG"  
/gene="novG"

CDS

5730. .7686

/note="putative regulatory protein"

gene

/codon\_start=1  
/transl\_table=11  
/product="NovG"  
/protein\_id="AAF67500.1"

CDS

7882. .9684  
/gene="novH"  
/gene="novH"

gene

/note="similar to peptide synthetase"

CDS

7882. .9684

gene

/codon\_start=1  
/transl\_table=11  
/product="NovH"  
/protein\_id="AAF67501.1"  
/db\_xref="GI:7688714"

CDS

7882. .9684

gene

/translation="WFNTRANKASDQSPPTPTESATLAEIWMERTVRSPSPAIWTNG  
ETLSYDEWNAENARLRLILLDEGAGPGRILVALALPRSHLIVISLAVAKAGAVFLPD  
VNHPRERLSYQLADARPAALLCTVRSAAARLPDGIEMPRVLDSPERTAVLDALPDIDL

CDS

7882. .9684

```

TDBRGGLAATLAVIYTSGRPKGVALTAGLALAAAKVAMKRVTCDSRVLQ
FASPGDRLVTELLAFTAGATLVFGTDTLAGDPLRLALDRGVSHVLRPRRSATM
SPDAPDLRVLVVAGACFAGLWAPGRLLINAYPTCTVCAATMTGPLEPTDEVT
IGRPIGVSVYIILADBRPAAFGIEGLYLSGALQAQYLNSPDILTAQMFVNPFAAD
GERMYRTGDLASRADIDILPHGRIDDOVELRGRFVELGEVESVLSQHPDVAQAAAL
WTPDASGPOLATVVVPAAGTTPSAGELRSHAGRFELDFMVPSAFTTIDAVPLTPGKT
DRAGLDPVKATPAGLGPRTPAEKVLCDIFERDLDFLVEIDVRSNFFENGNSILLVD
LLOAQAEGAGLMLPRTVIDHPTIEQLAAATLEB"
/gene=.10911
/genes="novi"
/9688..10911
/genes="novi"
/notes="cytochrome P-450"
/codon_start=1
/transl_table=11
/product="Novi"
/protein_id="AAFG7502.1"
/db_xref="GI:7688715"

Query Match 19.0%; Score 95; DB 1; Length 25617;
Best Local Similarity 54.0%; Pred. No. 0.00073; Mismatches 165; Indels 0; Gaps 0;
Matches 194; Conservative 0;

QY 140 TCAGTGTACCGGGGACGACGATACAGGATCCACTTCGCGACGCCGGGGAACCTCTGCT 199
DB 24019 TCACCCGCGGTTCCGGGTATGGGAACGATGAAATCTCGGACCCCGGAAGATATTTCCCTCT 23960
QY 200 CTTTCGCGATGATCTCGTCCGCGTGGTTCCAGGCGAAGAGCAGCGCTAGTCCACCGCGT 259
DB 23959 GCTTGTGATGATCTCCGGTCCGTTTCCAGGCGAAGATATTTCCCTCT 23900
QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGATGTGGTCCGGGGTGGCGGCGCT 319
DB 23899 GACCGCGCGCTGTTCCGGTACCGAGATCGGAATCTTGAGACCCCGGAAGATATTTCCCTCT 23840
QY 320 GCTTCGCGCGGTCTGTCGCACACCCAGGAGACAGGTCCGGACCGATGCCGAGAGT 379
DB 23839 GCTTCAGAAACGTTGTCGTCGAGAACCGGATCTCTGATGTCCGAGCCCGCAGCGG 23780
QY 380 TCGTCACGGTGGCGCTCTTCGCGTCCGCGTGGCTAGCCACACCGCTTGGCTCGGCT 439
DB 23779 CCAGCAGTGGCTGCGCTTGGTCCGTCGCGGTATCCCGCCACCGTCTTTCGCGTCCGCCA 23720
QY 440 TCAGCGAGTTCAGCAGGCGGACGAGTCCGTCGGATGCTCCGTCGCGACCGCTTCCGTACGATCGCGGAAC 498
DB 23719 CCACGCGCGTACGATGTTCCGGATCTGTGCGGACCGCTTCCGTACGATCGCGGAAC 23661

RESULT 31
AX707115/c 31248 bp DNA linear PAT 04-APR-2003
LOCUS
DEFINITION Sequence 1 from Patent WO03014352.
ACCESSION AX707115
VERSION AX707115.1 GI:29563420
KEYWORDS
SOURCE Streptomyces rishiriensis
ORGANISM Streptomyces rishiriensis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE
1
AUTHORS Heide, L. and Li, S.M.
TITLE Nucleic acids for aminocoumarin biosynthesis
JOURNAL Patent: WO 03014352-A 1 20-FEB-2003;
Universitaet Tuebingen (DE)
FEATURES
Location/Qualifiers
source 1..31248
/organism="Streptomyces rishiriensis"
/mol_type="unassigned DNA"
/db_xref="taxon:68264"

ORIGIN
Query Match 18.8%; Score 94; DB 6; Length 31248;
Best Local Similarity 53.9%; Pred. No. 0.00093;
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 140 TCAGTGTACCGGGGACGATACAGGATCCACTTCGCGACGCCGGGGAACCTCTGCT 199
DB 29686 TCACCGCGGTTCCGGGTATGGGAACATGAAACCGCGCGCCCTTCCAGAGAGATTTCT 29627
QY 200 CTTTCGCGATGATCTCGTCCGCGTGGTTCCAGGCGAAGAGCAGCGCTAGTCCACCGCGT 259
DB 29626 CTTTGTGATGATCTCCGACGCGTGTTCAGGCGCAACAGCAGTAGTAGTCCGCTACGT 29567
QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGATGTGGTCCGGGGTGGCGGCGCT 319
DB 29566 GGTGCTTGTGTTTCGCGGACCGATCGGAATCTGGAGCCCGGTAGGAGTTTCCCT 29507
QY 320 GCTTCGCGCGGTCTGTCGCACACCCAGGAGACAGGTCCGACCGATGCCCGCAGAGT 379
DB 29506 GCTTCAGGGTGTGTGTCGTCGAGAACCGGATCTCTGATGTCCCGACCGCGATCGCG 29447
QY 380 TCGTCACGGTGGCGCTCTTCGCGTCCGCGTGGCTAGCGCACACCGCTTGGCTCGGCT 439
DB 29446 TCAGCAGTGGCGCTTTCGCTGGTCCGCGCGGTATCCAGCACCTGTCTTCGCTCTGCCA 29387

RESULT 30
AX707121/c 25617 bp DNA linear PAT 04-APR-2003
LOCUS
DEFINITION Sequence 7 from Patent WO03014352.
ACCESSION AX707121
VERSION AX707121.1 GI:29563426
KEYWORDS
SOURCE Streptomyces caeruleus
ORGANISM Streptomyces caeruleus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE
1
AUTHORS Heide, L. and Li, S.M.
TITLE Nucleic acids for aminocoumarin biosynthesis
JOURNAL Patent: WO 03014352-A 7 20-FEB-2003;
Universitaet Tuebingen (DE)
FEATURES
Location/Qualifiers
source 1..25617
/organism="Streptomyces caeruleus"
/mol_type="unassigned DNA"
/db_xref="taxon:195949"

ORIGIN
Query Match 19.0%; Score 95; DB 6; Length 25617;
Best Local Similarity 54.0%; Pred. No. 0.00073;
Matches 194; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 140 TCAGTGTACCGGGGACGACGATACAGGATCCACTTCGCGACGCCGGGGAACCTCTGCT 199
DB 24019 TCACCCGCGGTTCCGGGTATGGGAACGATGAAATCTCGGACCCCGGAAGATATTTCCCTCT 23960
QY 200 CTTTCGCGATGATCTCGTCCGCGTGGTTCCAGGCGAAGAGCAGCGCTAGTCCACCGCGT 259
DB 23959 GCTTGTGATGATCTCCGGTCCGTTTCCAGGCGAAGATATTTCCCTCT 23900
QY 260 CGGGCGTGAACCGCTCCGGGTGGCCACCGGATGTGGTCCGGGGTGGCGGCGCT 319
DB 23899 GACCGCGCGCTGTTCCGGTACCGAGATCGGAATCTTGAGACCCCGGAAGATATTTCCCTCT 23840
QY 320 GCTTCGCGCGGTCTGTCGCACACCCAGGAGACAGGTCCGGACCGATGCCGAGAGT 379
DB 23839 GCTTCAGAAACGTTGTCGTCGAGAACCGGATCTCTGATGTCCGAGCCCGCAGCGG 23780
QY 380 TCGTCACGGTGGCGCTCTTCGCGTCCGCGTGGCTAGCCACACCGCTTGGCTCGGCT 439
DB 23779 CCAGCAGTGGCTGCGCTTGGTCCGTCGCGGTATCCCGCCACCGTCTTTCGCGTCCGCCA 23720
QY 440 TCAGCGAGTTCAGCAGGCGGACGAGTCCGTCGGATGCTCCGTCGCGACCGCTTCCGTACGATCGCGGAAC 498
DB 23719 CCACGCGCGTACGATGTTCCGGATCTGTGCGGACCGCTTCCGTACGATCGCGGAAC 23661
```

QY 440 TGAGCGAGTTACAGCGGAGCAGGTCGGTGGGATGTCCTCGAGTCGGCGGCGAA 497  
Db 29386 CCGAGTCGCGACCAAGTCCCGATGTCGGCGCACCGTTCGGTACCTCGCGGAA 29329

RESULT 32  
AF235050/c  
LOCUS Streptomyces rishiriensis strain DSM 40489 coumermycin A1  
DEFINITION biosynthetic gene cluster, complete sequence.  
ACCESSION AF235050  
VERSION AF235050.3 GI:38147033  
KEYWORDS Streptomyces rishiriensis  
SOURCE Streptomyces rishiriensis  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 35359)  
AUTHORS Wang,Z.-X., Li,S.-M. and Heide,L.  
TITLE Identification of the coumermycin A(1) biosynthetic gene cluster of Streptomyces rishiriensis DSM 40489  
JOURNAL Antimicrob. Agents Chemother. 44 (11), 3040-3048 (2000)  
MEDLINE 20493109  
PUBMED 11036020  
REFERENCE 2 (bases 1 to 35359)  
AUTHORS Wang,Z.-X., Li,S.-M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (16-FEB-2000) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany  
REFERENCE 3 (bases 1 to 35359)  
AUTHORS Wang,Z.-X., Li,S.-M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2000) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany  
REMARK Amino acid sequence updated by submitter  
REFERENCE 4 (bases 1 to 35359)  
AUTHORS Wang,Z.-X., Li,S.-M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (31-OCT-2000) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany  
REMARK Sequence update by submitter  
REFERENCE 5 (bases 1 to 35359)  
AUTHORS Wang,Z.-X., Li,S.-M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUL-2001) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany  
REMARK Sequence update by submitter  
REFERENCE 6 (bases 1 to 35359)  
AUTHORS Wang,Z.-X., Li,S.-M. and Heide,L.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUN-2002) Pharmazeutische biologie, Pharmazeutisches Institut, University of Tuebingen, Auf der Morgenstelle 8, Tuebingen 72076, Germany  
REMARK Sequence update by submitter  
COMMENT On Nov 3, 2003 this sequence version replaced gi:14625486.  
FEATURES  
source  
1..35359  
/organism="Streptomyces rishiriensis"  
/mol\_type="genomic DNA"  
/strain="DSM 40489"  
/db\_xref="taxon:68264"  
<1..1028  
/notes="ORF3"  
/codon\_start=3  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA11882.1"  
/db\_xref="GI:38147034"

CDS  
complement(1045..1536)  
/note="ORF2; similar to peptidylprolyl isomerase"  
/codon\_start=1  
/transl\_table=11  
/product="putative FK506-binding protein"  
/protein\_id="AA11883.1"  
/db\_xref="GI:38147035"  
/translation="MCPAGTASDGSDEAVMSPTPEVDVDPGDAPELTIIRDLIVG  
EGVEVKGVMVRYHYGVTFESGKEFDASWDGQPFKFGVSGKVIKWDGKRGVMV  
GGREIIVPRLGYNQSPSLIPPSSTLVFVVDLLDAYSSATGWSNAQRGHGVIQGR  
AEP"  
1535..1975  
/note="ORF1"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AA11884.1"  
/db\_xref="GI:38147036"  
/translation="MDTPCGRWGYRPSALSRPPTRRKGRCHSCRGPPVRVPTDG  
SRCPVCSARSTPAERADCSMPSPASPSPTRGESDAGWDESGQLWSVASGAA  
RRSWCMIRISQVORSASGRMLPCSGSGPAGAGRTMLSTS"  
2306..2959  
/gene="couB"  
2306..2959  
/gene="couB"  
/note="similar to NovE protein from Streptomyces  
sphaeroideus"  
/codon\_start=1  
/transl\_table=11  
/product="CouB"  
/protein\_id="AA11885.1"  
/db\_xref="GI:38147037"  
/translation="MVASGOTASKAGSGATPVRTAGDATPVDSRKSASHASCLGVEV  
SAESTRLQIPDLSLEAGRLGRLITVCDSSAWIGDWLFGONQYGDVRRAMEKT  
KLDQTLNANVAVARKEPSPRRDLTFQHHMEVASISOAQDHLPFAVLNKSRLQ  
LRKQIRASIGSEEDLRRHVQLSLQDLRLERWREARSGNLTLTDWILFVDEAL"  
4112..5071  
/gene="couG"  
/note="couG"  
4112..5071  
/gene="couG"  
/note="putative regulatory protein; CouG"  
/codon\_start=1  
/transl\_table=11  
/product="CouG"  
/protein\_id="AAG29778.1"  
/db\_xref="GI:11095212"  
/translation="WSNSGDEMTSLPKAIGKGVPIGSLPSPQQLRSRGSETEHI  
RALVETVELPPIVHRETRVIDGMRLRAAPRGDEWIEVVFIDASADVFAYVE  
LNRAGPLTMAERKSAARIMDSHPHRSKARTTGLAASTVASRSHSIATLER  
RTQDGTGTRPDGTEGRASSELLAQPNASVREVAKAISVGTASNARLRGEP  
AITALQAAITPRPASATGTGPDYSRLQSLRNPNPLRFNDVGRHLRLILUGAAGPS  
VEQIAQADSVPEHCRVWVIDMARECAAAWHLADRLAGRDTA"  
5408..5623  
/gene="couY"  
/note="couY"  
5408..5623  
/gene="couB"  
/note="similar to Mycobacterium tuberculosis Mbth"  
/codon\_start=1  
/transl\_table=11  
/product="CouY"  
/protein\_id="AAG29779.1"  
/db\_xref="GI:11095213"  
/translation="MATNPFDDENGVIYLLVNDGEGQSLPSPFALPKGVTILDEAS  
RQECIDYVNEHWTDMRPLISQPMGDE"

5929. 7728  
/gene="coulH"  
/note="cuncM"  
5929. 7728  
/gene="coulH"  
/note="coulH"  
/codon\_start=1  
/transl\_table=11  
/product="acyl-CoA synthetase"  
/protein\_id="AAG29780.1"  
/db\_xref="GI:11095214"  
/translation="MPSKGLNKASDRSATRTASATLAEIWEETVSRSSPAIVNS  
ETUSYDVARANQALRLLESEGGPGRVVALAPRSRMMVSVLAVAKAGAAFLPVD  
VNPEQRIAYMLDAEPAFLICTINTAVPKLPADIGVPPQLVLDSEATLADLPDIDV  
TEDEGSLAATLVIATVYTSSTGPKPGKAGLPLAALAAKAVTMQVTEDESRVLQ  
FASPSFATVTELIATVTSGLVLPVGSDDLVDGLAKVLRQGVSHAVLPAAVATM  
SPDAAPDLRVLAGECAPAGLEVQSPGRLLVNAVPTCTCAATWGLPTPTDVT  
IGRPVGSVHLLDELPAVAGEIGELVLSGAGLARGVLRNPDVTAHVFAVNPRAVL  
GERMYRGDLASRADDELPHGRIDQVZELGRFIELGESVLSQHPDVAQAVL  
RAGDDQQLAVVYPHTGTPAGELEHAGRLPDYMPVPSVYATIDALPVPVPSKT  
DRAALPKIKTRPAGQKPTSAEKLIFDIFDLFGJLGEIDVRSNFFEMGGNSILAVDL  
TORAQEAGLVLLPRTVIDHPTIEQLAAIANLGE"  
7732. 8955  
/gene="coulH"  
/note="cuncM"  
7732. 8955  
/gene="coulH"  
/note="coulH"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P-450"  
/protein\_id="AAG29781.1"  
/db\_xref="GI:11095215"  
/translation="MSTRPMVSADELDIDLASPIHAEYELDAVFRHLRANRPMYQ  
QPRGEQFWISYADYNEVYKDKAHTFTENGALALITLGGDSAGMALVTDGVR  
HQVNLISLGSFQPMLELTANSURETVGGLLALDRGEDDAARDISAKVPLGACD  
LLEIPQDREYLLGHTAHSTYADPEESWAKNEIILYFSKLKERRGGDRDM  
VSLLANRIGDHPILNAAEQIANCYGLMIGDETGRHAITGLVLIENPQWGLKNG  
DVLKATTEBALRWTPSLHGRKATGVVINGQIKAGEVSWVWSSANRDETVDA  
ADEKPLATPNKHTFAYGSHYCLHGLVGRMEVAVLDGLRLVGLDLEQIEERWLYS  
SILHGMSLPKITA"  
9115. 9891  
/gene="coulH"  
/note="cuncM"  
9115. 9891  
/gene="coulH"  
/note="coulH"  
/codon\_start=1  
/transl\_table=11  
/product="CouJ"  
/protein\_id="AAG29782.1"  
/db\_xref="GI:11095216"  
/translation="WTGPAASAVACEPVMVVTGAGRGIGATAERLAEGMVAIVVDR  
VEQDTATVSAIRAAGGRASGLGDVSAQAVTATVASAVEEFGHLDVLVNCAGITOD  
RLLLTMSDOEHLVDVNLGSMRCSFVGRHMEROGHRIINFSVVAARGNAGOTNY  
ATAKAIAGFTRTLAEPGPHGVTVNSIAPGFVATPMVAELTERLGLVDRDVLLEAM  
SSAVGRIQTPBEILAAVAVFVTRPESGYLTGETIIVEGRR"  
9888. 10625  
/gene="coulH"  
/note="cuncM"  
9888. 10625  
/gene="coulH"  
/note="coulH"  
Query Match 18.8%; Score 94; DB 1; Length 35359;  
Best Local Similarity 53.9%; Pred. No. 0.0009;  
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 140 TCACGTGTACGGGGGAGGTACAGGATCCACTGTCGCGCACGCCCGGGGAACTCCTGCT 199  
Db 33797 TCACGGCGGTTTCGGGTATGGGAACAATGAAACGGCGCCCTTCCAGGAACGACTTCT 33738  
QY 200 CTTTCCCATGATCTGTCGGGTGTTCCAGCGAAGAGCAGCGGTAGTCCACCGCT 259  
Db 33737 CTTTGTGTATGATCTCCGACGCGTAGTTCAGGCCAACAGCAGGTAGTCCGTACGT 33678  
QY 260 CGSGCGTGAACCGTCCGGGGTGGCCACCGGATGTGCGTGGGGGTGAGCCGCGCT 319  
Db 33677 GGTCCGTTGCTGTTCCGGCGACCATCGGAATCTCGAGCCCGTAGGAGTTTTCCT 33618  
QY 320 GCTTGGCCGGCGTGTGTGTCGACACCCAGAGACACAGGTCCCGACCGATGCGCAGAGT 379  
Db 33617 GCTTCAGGGCTGTGTGTGTCGTCGACACCGGATCTCTGATGTCCACGCCGATGCGG 33558  
QY 380 TCGTCACGTTGGCGCTTTCGCGCTGCGCGCTGACGCCACCGCTTGCCTCGGCT 439  
Db 33557 TCAGCAGTGGCTGCTTGTGTCGGTGGCGGTATCCAGCCACTGTCTTTCGGTCTGCCA 33498  
QY 440 TGACGAGTTACAGCGGGAGCAGGTGCGTGGGATGCCGTGCGAGTCCGCGGGCGAA 497  
Db 33497 CCAAGCTGGCAGCACTCCCGGATCTGTGCGCGCACCGCTTCCGTACGTCGCGGAA 33440  
RESULT 34  
AX574040/c

Db 33737 CTTTGGTGTATGATCTTCGACCGGTAGTTCAGGCGCAACAGCAGGTAGTTCGGTAGCT 33678  
QY 260 CGSGCGTGAACCGGTCGGGGTGGCCACCGGATGTGCGTGGGGGTGAGCCGCGCT 319  
Db 33677 GGTCCGTTGCTGTTCCGGCGACCATCGGAATCTCGAGCCCGTAGGAGTTTTCCT 33618  
QY 320 GCTTGGCCGGCGTGTGTGTCGACACCCAGAGACACAGGTCCCGACCGATGCGCAGAGT 379  
Db 33617 GCTTCAGGGCTGTGTGTGTCGTCGACACCGGATCTCTGATGTCCACGCCGATGCGG 33558  
QY 380 TCGTCACGTTGGCGCTTTCGCGCTGCGCGCTGACGCCACCGCTTGCCTCGGCT 439  
Db 33557 TCAGCAGTGGCTGCTTGTGTCGGTGGCGGTATCCAGCCACTGTCTTTCGGTCTGCCA 33498  
QY 440 TGACGAGTTACAGCGGGAGCAGGTGCGTGGGATGCCGTGCGAGTCCGCGGGCGAA 497  
Db 33497 CCAAGCTGGCAGCACTCCCGGATCTGTGCGCGCACCGCTTCCGTACGTCGCGGAA 33440  
RESULT 33  
AX707116/c  
LOCUS AX707116 35359 bp DNA linear PAT 04-APR-2003  
DEFINITION Sequence 2 from Patent WO03014352.  
ACCESSION AX707116  
VERSION AX707116.1 GI:29563421  
KEYWORDS Streptomyces rishiriensis  
SOURCE Streptomyces rishiriensis  
ORGANISM Streptomyces rishiriensis  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Heide, L. and Li, S. M.  
TITLE Nucleic acids for aminocoumarin biosynthesis  
JOURNAL Patent: WO 03014352-A 2 20-FEB-2003;  
Universitaet Tuebingen (DE)  
FEATURES  
source  
1. 35359  
/organism="Streptomyces rishiriensis"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:68264"  
ORIGIN  
Query Match 18.8%; Score 94; DB 6; Length 35359  
Best Local Similarity 53.9%; Pred. No. 0.0009;  
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 140 TCACGTGTACGGGGGAGGTACAGGATCCACTGTCGCGCACGCCCGGGGAACTCCTGCT 199  
Db 33797 TCACGGCGGTTTCGGGTATGGGAACAATGAAACGGCGCCCTTCCAGGAACGACTTCT 33738  
QY 200 CTTTCCCATGATCTGTCGGGTGTTCCAGCGAAGAGCAGCGGTAGTCCACCGCT 259  
Db 33737 CTTTGTGTATGATCTCCGACGCGTAGTTCAGGCCAACAGCAGGTAGTCCGTACGT 33678  
QY 260 CGSGCGTGAACCGTCCGGGGTGGCCACCGGATGTGCGTGGGGGTGAGCCGCGCT 319  
Db 33677 GGTCCGTTGCTGTTCCGGCGACCATCGGAATCTCGAGCCCGTAGGAGTTTTCCT 33618  
QY 320 GCTTGGCCGGCGTGTGTGTCGACACCCAGAGACACAGGTCCCGACCGATGCGCAGAGT 379  
Db 33617 GCTTCAGGGCTGTGTGTGTCGTCGACACCGGATCTCTGATGTCCACGCCGATGCGG 33558  
QY 380 TCGTCACGTTGGCGCTTTCGCGCTGCGCGCTGACGCCACCGCTTGCCTCGGCT 439  
Db 33557 TCAGCAGTGGCTGCTTGTGTCGGTGGCGGTATCCAGCCACTGTCTTTCGGTCTGCCA 33498  
QY 440 TGACGAGTTACAGCGGGAGCAGGTGCGTGGGATGCCGTGCGAGTCCGCGGGCGAA 497  
Db 33497 CCAAGCTGGCAGCACTCCCGGATCTGTGCGCGCACCGCTTCCGTACGTCGCGGAA 33440  
RESULT 34  
AX574040/c



LOCUS AX574040 1290 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 120 from Patent WO02079505.  
ACCESSION AX574040  
VERSION AX574040.1 GI:27551597  
KEYWORDS  
SOURCE  
ORGANISM  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
AUTHORS  
TITLE  
JOURNAL  
Patent: WO 02079505-A 120 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)  
FEATURES  
Location/Qualifiers  
1..1290  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
ORIGIN  
Query Match 18.5%; Score 92.4; DB 6; Length 1290;  
Best Local Similarity 53.8%; Pred. No. 0.0037; Indels 0; Gaps 0;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 142 ACCTGTACGGGGGACGTACAGGATCCACTGTCCGCCAGCCGGCGGAATCTCTGCTCC 201  
Db 1250 ACCCGAGGTGGCGGAGGAGATCAGTTGTTCCGGCCGCCACCGTGGCCGCTCG 1191  
QY 202 TTGCGCATGATCTCTGTCGGGGTGTTCAGGCGAAGACGAGCGGTAGTCCACCGGTGC 261  
Db 1190 CGGGCCACCATCTCTGTCGGAAGTGCACCGGACGACCATGAGTGTGTCGGGGGGGG 1131  
QY 262 GCGGTGAACGCGTCCGGGGTGCACACCGGATGTGTCGGGGGGGTGAGCGGCCCTGC 321  
Db 1130 GCCCGGAGTGGCCCTCGCTGATGATCTCGATGTCGATACCGAGGTTCCGCGCGGAC 1071  
QY 322 TTGGCGGGTGTGTCGACACCCAGGAGACCGTCCGACCGATGCGCGGAGTTC 381  
Db 1070 TTGTCGGGGTGTGCTCGCGGGGTACGGGATCAGCGTGGTTCGATCCCGCAGTACTG 1011  
QY 382 GTACAGTGGCGTCTTGGCGGTCGCGGTAGCCACACCGCTTGCCTCGGCTTG 441  
Db 1010 AGGAGGGTGTGCTTGGGAGGCGCGGTAGACGTCGACGGTGTTCGCTGTGCGC 951  
QY 442 AGCGAGTTCAGCAGGCGAGCAGGTGCGTGGGATGCGCTTCGACGTGCGCGGGAAC 499  
Db 950 AGACCATGCAGCATCTTACGAGTTGTCGCGGTGCGCGCGCGCAGCGTTGTCGCGAAC 893  
RESULT 35  
AX204973/c  
LOCUS AX204973 14252 bp DNA linear PAT 30-AUG-2001  
DEFINITION Sequence 8 from Patent WO015180.  
ACCESSION AX204973  
VERSION AX204973.1 GI:15394253  
KEYWORDS  
SOURCE  
ORGANISM  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Staffa, A., Zazopoulos, E., Mercure, S. and Nowacki, P.  
AUTHORS  
TITLE  
JOURNAL  
Patent: WO 015180-A 8 02-AUG-2001;  
Ecopia Biosciences Inc. (CA); Farnet, Chris (CA)  
FEATURES  
Location/Qualifiers  
1..14252  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"  
misc\_feature 459..1280

/note="ORF 5 (positive strandedness)"  
1280..2566  
/note="ORF 6 (positive strandedness)"  
2677..3747  
/note="ORF 7 (positive strandedness)"  
3899..4774  
/note="ORF 8 (positive strandedness)"  
4893..5303  
/note="ORF 9 (positive strandedness)"  
5385..6306  
/note="ORF 10 (negative strandedness)"  
6350..7204  
/note="ORF 11 (negative strandedness)"  
7371..8198  
/note="ORF 12 (negative strandedness)"  
8304..9098  
/note="ORF 13 (negative strandedness)"  
9462..10493  
/note="ORF 14 (positive strandedness)"  
10665..11384  
/note="ORF 15 (negative strandedness)"  
11387..12700  
/note="ORF 16 (negative strandedness)"  
12971..14185  
/note="ORF 17 (negative strandedness)"  
ORIGIN  
Query Match 18.5%; Score 92.4; DB 6; Length 14252;  
Best Local Similarity 53.6%; Pred. No. 0.0019;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 142 ACCTGTACGGGGGACGTACAGGATCCACTGTCCGCCAGCCGGCGGAATCTCTGCTCC 201  
Db 2529 ACCCGAGGTGGCGGAGGAGATCAGTTGTTCCGGCCGCCACCGTGGCGGCTCG 2470  
QY 202 TTGCGCATGATCTCTGCGGGTGTTCAGCGAAGACGAGCGGTAGTCCACCGGCTGC 261  
Db 2469 CGCGCCAGATCTCTGTCGGAAGTGCACCGGACGACCATGAGTGTTCGCGCGCGG 2410  
QY 262 GCGGTGAACGCGTCCGGGGTGCACACCGGATGTGTCGGGGGTGAGCGGCGCTGC 321  
Db 2409 GCCCGGAGTGGCGCTCGCTGATGATCTCGATGTCGATCGGAGGTCGCGGCGCGG 2350  
QY 322 TTGCGCGCGTGTGTCGACACCCAGGAGACGAGTCCGACCGATGCGCGGCGGAAGTTC 381  
Db 2349 TTGTCGGGTGCGCTCGCGGCGGTACGGATCAGCGTGGGTGATCCCGCAGTACTGC 2290  
QY 382 GTACAGTGGCGCTTTCGCGGTGCGCGGTACGCCACACCGCTTGGCGCTTGGCGCTTG 441  
Db 2289 AGGAGGGTGTGCGCTTGTGAGGGCGCGGTAGACGTCGACGGTGTCTTCGCTGCGGC 2230  
QY 442 AGCGAGTTCAGCAGGCGGACGAGTTCGTCGGGATGCGCTTCGACGTGCGGCGGAAC 499  
Db 2229 AGACCATGCAGCATCTTACGAGTTGTCGCGGTGCGCGCGCGCAGCTTGTTCGCGAAC 2172  
RESULT 36  
AX574200/c  
LOCUS AX574200 48221 bp DNA linear PAT 07-JAN-2003  
DEFINITION Sequence 280 from Patent WO02079505.  
ACCESSION AX574200  
VERSION AX574200.1 GI:27551679  
KEYWORDS  
SOURCE  
ORGANISM  
Micromonospora carbonacea  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Micromonosporineae; Micromonosporaceae; Micromonospora.  
REFERENCE  
1 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
AUTHORS  
TITLE  
JOURNAL  
Patent: WO 02079505-A 280 10-OCT-2002;  
Ecopia Biosciences Inc. (CA)

FEATURES	Location/Qualifiers	
	1..48221	
	/organism="Micromonospora carbonacea"	
	/mol_type="unassigned DNA"	
source	/db_xref="taxon:47853"	
ORIGIN		
	Query Match	
	Best Local Similarity 53.6%; Score 92.4; DB 6; Length 48221;	
	Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;	
QY	142	ACGTGTACGGGGGAGCTACAGATCCATCTCCGCGACCGCGGGAACCTCTCTCC 201
Db	8244	ACCCGCGAGCTGGCGACGCGGAAGATCAGCTTGGTTCGGCGCGCCACCGTGGCGGCTCG 8185
QY	202	TTGCCCATGATCTCGTCGGCTGTTCAGCGCGAAGACGCGGTAGTCCACCGCGTCG 261
Db	8184	CGCGCCAGCATCTCGTGTGAAAGTGCACCGGACGACGAGTGTAGTGTCTCCGCGCGCGG 8125
QY	262	GGCTGAACGGCTCGGGGTCCGACCGGATGTGCTGCGGGGGTGAACCGGCTCTGC 321
Db	8124	CGCCGCGAGTCCGGCTCGCTGATGATCTCGATGTCCGATCCGAGGGTCCCGCGCGGAC 8065
QY	322	TTGCCCGCGCTGTGTCACACCGACGAGACGAGTCCGACCGATGCGCGAGAGTTTC 381
Db	8064	TTGTCGGGTTGCGCTCGCGCGCTGACGGATCAGCGTGGTGTGATCCCGGAGTCTGC 8005
QY	382	GTACAGGTGGCGCTTTCCCGCTCGCGCGCTGACCGACCGACCGCTGCGGCTTCG 441
Db	8004	AGGAGGTTGTCCTTCTGTCGAGCGCGCTGAGAGTGTGCTGCGCTTCGCGC 7945
QY	442	AGCGAGTTCAGACGGCGGACGAGTTCGCTGCGGATGCGCTCGCGCGGCGAAC 499
Db	7944	AGACCATGAGCATCTTACGAGTTCGTCGCGGTGCGCGCGGCTGTCGCGGAC 7887
RESULT 37	RME603646/c	
LOCUS	303100 bp DNA linear BCT 05-JUL-2002	
DEFINITION	Rhizobium meliloti (Sinorhizobium meliloti) 1021 complete psymb; segment 5/6.	
ACCESSION	AL603646 AL591985	
VERSION	AL603646.1 GI:15140973	
KEYWORDS	Sinorhizobium meliloti (Rhizobium meliloti)	
SOURCE	Sinorhizobium meliloti	
ORGANISM	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.	
REFERENCE	1 Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorholter, F.J., Hernandez-Lucas, I., Becker, A., Cowie, A., Gouzy, J., Golding, B. and Puhler, A. The complete sequence of the 1,683-kb psymb megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9889-9894 (2001)	
AUTHORS	21396508	
TITLE	11481431	
JOURNAL	epub ahead of print	
MEDLINE	2 (bases 1 to 303100)	
PUBMED	Weidner, S.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (07-JUN-2001) Weidner S., Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany	
TITLE	Submitted on behalf of Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany and Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4K1 Canada	
JOURNAL	mailto:Stefan.Weidner@Genetik.Uni-Bielefeld.DE pEXO, psymb.	
COMMENT		
FEATURES	Location/Qualifiers	
	1..303100	
source	/organism="Sinorhizobium meliloti"	
/mol_type="genomic DNA"		
/strain="1021"		
/db_xref="taxon:382"		
/plasmid="pSymb"		
gene	complement(167..967)	
CDS	/gene="thiD OR Smb20962"	
	complement(167..967)	
	/gene="thiD OR Smb20962"	
	/EC_number="2.7.4.7"	
	/function="Small Molecule Metabolism; Biosynthesis of cofactors, carriers; thiamin"	
	/note="Product confidence: probable"	
	gene name confidence: putative"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
/product="probable phosphomethylpyrimidine kinase protein"		
/protein_id="CAC49487.1"		
/db_xref="GI:15140974"		
/db_xref="GOA:P56904"		
/db_xref="SWISS-PROT:P56904"		
/translation="MTAIALSIAGSDGGGAGIQADLKTFALSALGVYGASVITAITQAN TEGTVADVDAEIVSAQMDAVFDLDVKAVIGMVSRRRTIAAIDGLRRFGKRAVV DMVATGSDALLRPDAVALIELLPLAVLTPNLAAALMTGRAJAGBAEMARQA EALMTGHAHLVKGCHLKGAEATDLPDGDTLVRLPAGRIETENDHGTCTLSAALA AIAKGVPLIEAVSAKAYLHAALISAADRIEIGCGRGPVHFHFRWKD"		
1002..1307		
gene	/gene="Smb20963"	
CDS	1002..1307	
	/gene="Smb20963"	
	/function="MISCELLANEOUS; Not classified regulator"	
	/note="Product confidence: putative"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative transcriptional regulator protein"	
	/protein_id="CAC49488.1"	
	/db_xref="GI:15140975"	
/db_xref="GOA:Q92UN5"		
/db_xref="SPTREMBL:Q92UN5"		
/translation="NRVVRQMTSSIDLTQKHGDPGKDEKANLPLAAIPFLRRHDP DOVGSGSTPSQRLVLPSECAIYAARQARAKACKTYLLGQPAKIYPIQGLK"		
1412..2071		
gene	/gene="Smb20964"	
CDS	1412..2071	
	/gene="Smb20964"	
	/function="CELL PROCESSES; Protection responses"	
	/note="Product confidence: putative"	
	gene name confidence: hypothetical"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative anti-oxidant protein, AhpCTSA family"	
	/protein_id="CAC49489.1"	
/db_xref="GI:15140976"		
/db_xref="GOA:Q9R9R2"		
/db_xref="SPTREMBL:Q9R9R2"		
/translation="MSLRINDTAPDTAETGTTNFHWDGHWAVLFSPHKNTPT CTTELGMAGIEPEFRKGVKIGISVDESHSKWMDIKVATGEVDYFLIGLRDRI KVAKLYMLPAGAGTSEGTADNATSVTVIGPKKIKILITPYMTTGNFNEIRI RAIDSIQLTAKHOVATPANWQQEDVIITAAVSNEDAVQRFSGSDTLVLYLKTKOP A"		
2189..3187		
gene	/gene="Smb20965"	
CDS	2189..3187	
	/gene="Smb20965"	
	/function="MISCELLANEOUS"	
	/note="Product confidence: putative"	
	gene name confidence: hypothetical"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative exported enzyme, similar to	

FEATURES	Location/Qualifiers	
	1..1021	
source	/db_xref="taxon:382"	
	/plasmid="psymb"	
	complement(167..967)	
	/gene="thiD OR Smb20962"	
	complement(167..967)	
	/gene="thiD OR Smb20962"	
	/EC number="2.7.4.7"	
	/function="Small Molecule Metabolism; Biosynthesis of cofactors, carriers, thiamin"	
	/note="Product confidence : probable"	
	Gene name confidence : putative"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="probable phosphomethylpyrimidine kinase protein"	
	/protein_id="CAC49487.1"	
	/db_xref="GI:15140974"	
	/db_xref="GOA:P56904"	
	/db_xref="SWISS-PROT:P56904"	
	/translation="MTAIALSIAGSDSGGAGIADLTKFSALGVYGVASVITAITAQN TRGVTAEDVSABIVSQMDAVDFDLVDYKAVKIGMVSRETIAADLGRFGKRAVY DPMVATGDALLRPDAVALIEELLPLAVITPFLAEALMTGRAIAGDAEWAQOA EAIMRTGAHLVKGHLKGEATDLFPDGTILVKLPAGRIETRNHGTGCTLSAIA AGLAKGVPLIEAVSAKAYLHAAISADRLTEIGQGRGFVHFHFWKD"	
	1002..1307	
	/gene="Smb20963"	
	1002..1307	
	/gene="Smb20963"	
	/function="MISCELLANEOUS; Not classified regulator"	
	/note="Product confidence : putative"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative transcriptional regulator protein"	
	/protein_id="CAC49488.1"	
	/db_xref="GI:15140975"	
	/db_xref="GOA:Q92UN5"	
	/db_xref="SPTREMBL:Q92UN5"	
	/translation="MRVVRQMTSSIDLQTKHGDPKGEKANLPLAIFPFLRRHDP DOVGSGSTPSQHLVLPSECAIYAAVAERCQARAKACKTYLLOQPAKIYPIQKL"	
	1412..2071	
	/gene="Smb20964"	
	1412..2071	
	/gene="Smb20964"	
	/function="CELL PROCESSES; Protection responses"	
	/note="Product confidence : putative"	
	Gene name confidence : hypothetical"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative anti-oxidant protein, AhpCTSA family"	
	/protein_id="CAC49489.1"	
	/db_xref="GI:15140976"	
	/db_xref="GOA:Q9R9R2"	
	/db_xref="SPTREMBL:Q9R9R2"	
	/translation="MSLRINDTAPDTAETTQGTINFEWIGDWAVLFSPKXFTFV CTTGLGAMAGIEPEFRKRGVKIGISDVESHKWKNDIKVATGEVDYPLIGDRDL KVAKIXDMLPAGAGETSEGTTPADNATVRSYVIGPKKIKLITVMTGRNFNELL RAIDSITQLTAKHQVATPANWQGEDVIITAASVEDAVORFGSFDVILPYLTKQPT A"	
	complement(2189..3187)	
	/gene="Smb20965"	
	complement(2189..3187)	
	/gene="Smb20965"	
	/function="MISCELLANEOUS"	
	/note="Product confidence : putative"	
	Gene name confidence : hypothetical"	
	/codon_start=1	
	/evidence=not experimental	
	/transl_table=11	
	/product="putative exported enzyme, similar to"	

FEATURES	Location/Qualifiers	
	1..303100	
source	/organism="Sinorhizobium meliloti"	

```
/protein_id="CAC49492.1"
/db_xref="GI:15140977"
/db_xref="GOA:Q92UN4"
/db_xref="SPTREMBL:Q92UN4"
/translation="MLVITLALLGFHSAQAQDSVTTILTRLERELKPLKVVV
ADGSTIAEGRGHTPESSTNIKASKSIYSVALVGIADKGLDGGPKIAPLKED
LPSQADRIINDITIKLLSMQTLAGEKSPNRYVSRNWRVFAALQPFDDPEGRM
LYSTATLLSLIITKGVGKPTALAREWLDVEGFRICAWERDFQGIYLGNGQVMS
ARSLAFGEILYRNRTGRQVDPADWVELLSWRPTASRTFGDGYGYGWFTRIGDE
DYVYANGYGGQMLYTVPSGLNLTVMVMTSDSGPSARNGYRDALHGLVLAIIISTAKAD"
complement(3296..5236)
/gene="lacZ2 OR Smb20966"
complement(3296..5236)
/gene="lacZ2 OR Smb20966"
/EC_number="3.2.1.23"
/function="Small Molecule Metabolism; Degradation; carbon
compounds"
/notes="Product confidence : putative"
Gene name confidence : putative
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative beta-galactosidase protein"
/protein_id="CAC49491.1"
/db_xref="GI:15140978"
/db_xref="GOA:Q92UN3"
/db_xref="SPTREMBL:Q92UN3"
/translation="MCYIEPQMRGRWAEDARMVELGLSWRIGEFPAWKIEPRSGE
FWELMDEAIDVLGKAGLVILGTPAAPKVLNRYPEILPVDATGAVRKFGRHHY
CFSSRYSEARAIITAMARYGEHIYHWAOITDNEYGDHDTIYSABAVGAPRLWL
AERYGSDLELNAWGTFSWMSRYDPEEIDLPNNLVEBPSPTHGYDFIRFSDQVKSF
NKAQVDIIRAHSGRPVTHNFSQNTDFPHRVGDEDIDIASMDVYFGLGNGLAALK
DREHVLKVDQDPAHPHDLVRAVGRGVVVMEOQPPVNWAAHQSPADGWLRLWTW
LAYAHGVDMVSVFRWQAPPAEQFHAGLLPNSADQGYLEAVRVAEMKLEGEV
RGAQVAIVLDFESRWATVLFQGRSISFASVALDWSYLELPGVDVDFICGHSDIAG
YKILAPLDVIAEAEVERLITRADKAVFGARSKTRDMHPIPELGGPGLAKLIDIS
VRSYLVAFHSEETVLYGNEAYEAGCWRETIVRTSETVIASFDEYRNAPALVGSDKA
VILVAVANGALLDKVIGDALGWAGLESPLDGLRVTNRGNLRFAPNFGRIPEVPAG
ASAEFYVGGRTLKPVDAVIRE"
complement(5426..6460)
/gene="Smb20967"
complement(5426..6460)
/gene="Smb20967"
/function="MISCELLANEOUS; Not classified regulator"
/notes="Product confidence : putative"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transcriptional regulator, lacI family
protein"
/protein_id="CAC49492.1"
/db_xref="GI:15140979"
/db_xref="GOA:Q92UN2"
/db_xref="SPTREMBL:Q92UN2"
/translation="MORPIKTIARETGLSIATVSKALKHSPQVRPEPRAVVLDAER
VGYELNVQLRTGTYQVAALMTAPGRNEWEVGEYIAQLLSGWSALESPRVS
LYAVRDFESQETISQIVSKADGIIISGRADDDRIELMQEAFYFVYIGMSIHNH
PHAYVADMEQMLRLSVARLIERHRIALLNPKQFAYGVVRLSRYKALEKAGLPY
DPALVAGSLTPAFGRNVVMTSALADPTTAVICANEATLGAFGSGFARGLVHGRDA
VINATDNLNVSOYFAPPTITSYLYPIGEPSSALLGKFTLRMEGEPDPAQLTLMPDLIE
RSDRLRPER"
complement(6468..8180)
/gene="Smb20968"
complement(6468..8180)
/gene="Smb20968"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical"
Gene name confidence : hypothetical
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
```

```
/protein_id="CAC49493.1"
/db_xref="GI:15140980"
/db_xref="GOA:Q92UN1"
/db_xref="SPTREMBL:Q92UN1"
/translation="MLPFDIDHVPFSRKGRFLTSTWRVGRDGERALYLRRVSGGDER
PSLGRCLVEFLDGEGRPAATRFQLSPDQIVARTGRIYVRFVIGDGERLHIREAGV
RLVLESRDYVYRTTGGSCCLVAAGENVKEIPIRGELAVSGKWDORSTDVSTFL
SGSGTTEGVNDVFPRLPPPEGSGFEAAHAAVTAEGFANYSIPAGVFGQEEAQLAA
YLLWANTVFAEGVITTPALPYMSKNGMINIWSMDNFSALGVAAFPDESLAFQGAALFD
HODASGLLDYNDREALFAFTKPPVHVWAVSWMARENFAFLTPERRAYLRDAIGRQV
SYMLTGRAGDNLVPSYFNGDSGMDNASFFABGGPVLSPDLFPVLIILACEALANLLE
DPPGKAARLRIADLEQALLVGLTGTETFAARLAADPERILFGBESLIOPMPLLLASR
LPMACRQIVARIVEGAFITDMGPATESPRSPVEDDGYWRGPIWAPTVLLWDGLRR
QEGALAREIAQKFCALANAHGVAENFDARSGRGLDRAPATISAVYLLLAQSLRDEK
PVKS"
complement(8184..8996)
/gene="Smb20969"
complement(8184..8996)
/gene="Smb20969"

Query Match 18.2%; Score 90.8; DB 1; Length 303100;
Best Local Similarity 53.4%; Pred. No. 0.0013;
Matches 191; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 140 TCAGTGTACGGCGGGGACGTACAGGATCCACTGTCCGCCAGCCCGCGGAATCTTGGCT 199
DB 158066 TCACCGCACCTTCGGAATCGCGACGACGAGCGCGCCGCCAGGCGGATATAGCGCT 158007

QY 200 CTTTCGCATGATCTCTTCGCGCTGTTCCAGGCGAGACGAGCGGTAGCTCCACCGCT 259
DB 158006 GGGCGGCGAGCATCTCTCGGTTCAGGTCAGGCGGAGATGACACATAGTCGGGTTTA 157947

QY 260 CGGCGGTGAACCGCTCCCGGGTGGCGACCGGGATGTGCGTCCGCGGGGTGAGCGCGCCT 319
DB 157946 CGGCTCGATCTCTCGCGGATCGTAGATCGAATGTGCTCGCGGCGAGAGCTTGCCT 157887

QY 320 GTTTCGCGGGTCTGTGTCACACCGAGACAGGTCTCGGACCATCGCCAGAGT 379
DB 157886 GTTTCAGATCGTTCGCGTCCGACGATGAAGTCGATGTCGACGCGGTGAGCGCGGACAGCT 157827

QY 380 TCGTCACGTCGCGCTCTTTCGCGCTCGCGCTACGCCACACCGCTTGCCTTCGCGCT 439
DB 157826 TCAGGAGGTGTTGCCCTTTTCCCGCGGCCATAGCGCGCATGCTTGTTCGCTTC 157767

QY 440 TGAAGCGAGTTCAAGCGGCGAGAGGTTCGGTTCGGATGTCCTCGACGTCCGCGCGGCGAA 497
DB 157766 TGGCGTCCGCGAGAACGCGCGGAATCCATCGCAGACCGAAGCGATCGCTTGCAGAA 157709

RESULT 38
AX574197/c
LOCUS AX574197 45055 bp DNA linear PAT 07-JAN-2003
DEFINITION Sequence 277 from Patent WO02079505.
ACCESSION AX574197
VERSION AX574197.1 GI:27551676
KEYWORDS Streptomyces mobaraensis
SOURCE Streptomyces mobaraensis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Farnet,C.M., Zazopoulos,E. and Staffa,A.
TITLE Compositions and methods for identifying and distinguishing
orthomycin biosynthetic loci
JOURNAL Patent: WO 02079505-A 277 10-OCT-2002;
Ecopia Biosciences Inc. (CA)
FEATURES
source 1..45055
/organism="Streptomyces mobaraensis"
/mol_type="unassigned DNA"
/db_xref="taxon:35621"

ORIGIN
```

Query Match 18.1%; Score 90.6; DB 6; Length 45055;  
 Best Local Similarity 52.5%; Pred. No. 0.0023;  
 Matches 198; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 122 CACCGCGGGTACGTCGTCAGTGTACGCGGGGACCTACAGGATCCACTGTCCGCCAG 181  
 DB 9974 CCCCCGACCGGTGTACACCATTTCCAGGGCGGACGCGGAAGACAGGTTCGCCCGG 9915  
 QY 182 CCGCGCGGAATCTCTCTCTTCCGCGATGATCTCTCGCGGTGTCTCCAGCGAAGACA 241  
 DB 9914 CCGCCAGGAATCTCTCTCTTCCGCGGACCATCGCTCCCGGTAGATCCACGCGAGGACCA 9855  
 QY 242 GCGGTGTATCCACCGGTCTCGGCGGTGAACGCTCGGGGTGCGACCGGGATGCGTGC 301  
 DB 9854 GGAACCGGTCTCGGCTCCAGCGCCCGGCTCTCTTCGGAACGATCGGATGTCGTTGC 9795  
 QY 302 CGGGGTGTAGCGCGCCCTCTCTTGGCGCGGTCTGTGTGACACCGAGAGACCGAGTCCG 361  
 DB 9794 CGGGGTGTAGCGCGCCCTCTCTTGGCGCGGTCTGTGTGACACCGAGTCCGATGTCGTTGC 9735  
 QY 362 GACCGATGCGCGAGGATTCGTACGCTGGCGCTCTTCCGCGTCCGCGGTACCGCACCA 421  
 DB 9734 CGTCCAGCGCGAGTACTGGAGATGACGTTGCCCTTGTGGAGCGCCGTTACCGAGGG 9675  
 QY 422 CCGGTGTCCTCGCGCTTTCAGCGAGTTCAGCGGGGAGGAGTCTGTGCGGATGCGCTT 481  
 DB 9674 TCGCGAGCGCTTGTCTCCGCGAGTGTGTGAGGAGTCCACGAGCGGTGCGGGATTCCT 9615  
 QY 482 CGAGCTCGCGCGGAC 498  
 DB 9614 CGGTCCGCGCGGAC 9598

RESULT 39  
 AX574038/c 1224 bp DNA linear PAT 07-JAN-2003  
 LOCUS AX574038  
 DEFINITION Sequence 118 from Patent WO02079505.  
 ACCESSION AX574038  
 VERSION AX574038.1 GI:27551596  
 KEYWORDS  
 SOURCE Streptomyces mobaraensis  
 ORGANISM Streptomyces mobaraensis  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomycetes.

REFERENCE 1  
 Farnet, C.M., Zazopoulos, E. and Staffa, A.  
 Compositions and methods for identifying and distinguishing  
 orthomyxovirus biosynthetic loci  
 Patent: WO 02079505-A 118 10-OCT-2002;  
 Ecopia Biosciences Inc. (CA)  
 Location/Qualifiers  
 1..1224  
 /organism="Streptomyces mobaraensis"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:35621"

FEATURES  
 source

ORIGIN

Query Match 17.6%; Score 87.8; DB 6; Length 1224;  
 Best Local Similarity 52.6%; Pred. No. 0.015;  
 Matches 191; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 136 CTGCTCACGTGTACCGCGGGACGTACAGATCCACTGTCGCCGACCGCGCGAATCC 195  
 DB 1223 CACACCATTTCCAGGGCGGAGCGGAGACAGGTTCGCCGCGCGCGCGCGAAGTCC 1164  
 QY 196 TGTCTCTTCGATATCTCTGTCGCGGTGTTCAGGCGAAGAGAGCGCGTGTCTCCACC 255  
 DB 1163 CGTTCGCGGCGACCATCTGCTCCGCGTAGATCCACGCGGAGACCGAGACCGTCTCCGC 1104  
 QY 256 GGTCTCGCGGTGAACGCGTCCGCGGTGCGACCGGATGTCGTCGCGGGGTGACGCGG 315  
 DB 1103 TCCAGCGCGCGGCTCTCTTCGAGAGATCGGATGTCGTCGCGCGGCGTGAAGAG 1044

QY 316 CCTCTTGGCGCGGTCTGTGTGTCACACCCAGGAGACAGGTTCGGAGCATGCCGCGAG 375  
 DB 1043 CGGAACCTTGTCTCTCTTCACTTCGCGGATGCAACGCGAGGAGCTCTGTCAGGCGCGAG 984  
 QY 376 AGATTCTGTACCGGTGCGCTCTTGTGCGGTGCGCGGTACGCGACACCGCTTCCCTCG 435  
 DB 983 TACTTGAGGATGACCTTCTGCTTGTGAGGCGCGGTACCGGAGGTGCGGAGGCTTGTG 924  
 QY 436 GCTTTGAGCGAGTTCTCAGCAGGCGCGAGGTGCGTTCGCGATGCCCTTCGAGCTCGCGCGCG 495  
 DB 923 TCCCGCGAGGTGTTGAGGAAGTCCACGAGCGGTGCGCGGATTCCTCGTCCGCGCGCGCG 864  
 QY 496 AAC 498  
 DB 863 AAC 861

RESULT 40  
 AB088224 210614 bp DNA linear BCT 11-JUN-2003  
 LOCUS Streptomyces rochei plasmid pSLA2-L DNA, complete sequence.  
 ACCESSION AB088224  
 VERSION AB088224.1 GI:30698345  
 KEYWORDS  
 SOURCE Streptomyces rochei  
 ORGANISM Streptomyces rochei  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycineae; Streptomycetaceae; Streptomycetes.

REFERENCE 1  
 Kinashi, H., Fujii, S., Hatani, A., Kurokawa, T. and Shinkawa, H.  
 Physical mapping of the linear plasmid pSLA2-L and localization of  
 the eryA and actI homologs  
 Biosci. Biotech. Biochem. 62, 1892-1897 (1998)  
 2  
 Suwa, M., Sugino, H., Sasaoka, A., Mori, E., Fujii, S., Shinkawa, H.,  
 Nimi, O. and Kinashi, H.  
 Identification of two polyketide synthase gene clusters on the  
 linear plasmid pSLA2-L in Streptomyces rochei  
 Gene 246 (1-2), 123-131 (2000)  
 20231737  
 10767533

REFERENCE 3  
 Hiratsu, K., Mochizuki, S. and Kinashi, H.  
 Cloning and analysis of the replication origin and the telomeres of  
 the large linear plasmid pSLA2-L in Streptomyces rochei  
 Mol. Genet. Genet. 263 (6), 1015-1021 (2000)  
 20408175  
 10954087

REFERENCE 4  
 Mochizuki, S., Hiratsu, K., Suwa, M., Ishii, T., Sugino, F., Yamada, K.  
 and Kinashi, H.  
 The large linear plasmid pSLA2-L of Streptomyces rochei has an  
 unusually condensed gene organization for secondary metabolism  
 Mol. Microbiol. 48 (6), 1501-1510 (2003)  
 22676886  
 12791134

REFERENCE 5 (bases 1 to 210614)  
 Mochizuki, S., Hiratsu, K. and Kinashi, H.  
 Direct Submission  
 Submitted (15-JUL-2002) Haruyasu Kinashi, Hiroshima University,  
 Department of Molecular Biotechnology, Graduate School of Advanced  
 Sciences of Matter, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima  
 739-8530, Japan (E-mail: Kinashi@hiroshima-u.ac.jp)  
 Tel:81-824-24-7869, Fax:81-824-24-7869  
 The nucleotide sequence has been determined by using restriction  
 fragments and nested deletion fragments of the ordered cosmid  
 library of pSLA2-L.  
 pSLA2-L is 210,614 bp long with G+C content of 72.8 % and carries  
 143 ORFs.  
 Gene prediction was based on the unique codon usage in Streptomyces  
 (Bibb et al., Gene 30:157-66 (1984)) using the FramePlot program of  
 Ishikawa and Hotta (FEMS Microbiol Lett 174: 251-253 (1999)) as  
 implemented at

<http://www.nih.gov/cgi-bin/frameplot.pl>. Where possible we chose an initiation codon (atg, gtg, ctg or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13 bp before the initiation codon). If this could not be identified we chose the most upstream initiation codon.

Three-fourths of the pSLA2-L DNA are occupied by biosynthetic genes (three PKS gene clusters for lankacidin, lankamycin and an unknown type II polyketide, a carotenoid biosynthetic gene cluster, many regulatory genes and others). The range of each biosynthetic gene cluster has been deduced by comparison with similar gene clusters in most cases and may be revised in future.

## FEATURES

## source

```
Location/Qualifiers
1..210614
/organism="Streptomyces rochei"
/mol_type="genomic DNA"
/strain="7434AN4"
/db_xref="taxon:1928"
/plasmid="pSLA2-L"
/notes="linear plasmid"
1..1992
/notes="left terminal inverted repeat, TIR-L; shows 99.4 %
(1981/1992) sequence identity to TIR-R (complement
(208623..210614))"
683..2188
/notes="N-terminal sequence is almost identical (435/437)
with that of ORF143 at the right end of pSLA2-L until the
inner end of TIR-L
ORF1 (501 aa)
similar to AL590463 Streptomyces coelicolor putative
helicase, SCP1.136 (879 aa); homology is seen until the
inner end of TIR-L"
/codon_start=1
/transl_table=11
/product="putative helicase"
/protein_id="BAC76459.1"
/db_xref="GI:30698346"
```

## LTR

## CDS

```
/translation="MSTGKTDQREAAQOEAVDVVRALFVRLIAPRGLRTQVIM
ATGSGTTRVAARSAELRAGRLVLPVSLDLTQTEAAWEAGRTGPMIGVSLRGD
VAPNTDVEELVDVVRPEKVTATYASLGILTEARHGGELPCMDLIVYDEAHT
SGRLGFWAHDHNTIRPISRLRYMTATRLWQLDDEACAGELVSMEDDPDLFG
ARCTTLTSEADRGICAPYVVCVDITDTQQAQLLGVGRSDEVRGARLAALQTA
LKASSEENFRFTLVPHVMKEAFAAGLPDVAKGLHAAGPGLYPRITWANNLCGEH
KPHRRVVEFTSGIATDGTVVEKGLSVKVLGEGVTRCDLSYVADVRGSPDL
VQAVGRALRTQPGQKVASLVPVLLDPCGETADNMLTSRPNGLALLPQAPLLTGHG
EEGFRAARPGREVRVILGVGRGAQASGPASRITPSSMSLSISNEAP"
complement(2315..3595)
/notes="ORF2 (426 aa)
similar to AE004736-10 Pseudomonas aeruginosa hypothetical
protein (442 aa)"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="BAC76460.1"
/db_xref="GI:30698347"
```

## CDS

```
/translation="MRKPKALLGIPLLILVAVLAPAAHSAADTYVTWAPPTAA
PADATVYEDQLRQVHLSVAGDSLRVTFNEFGTSLTIGEVHAARPAAGPATAV
DQGTDRVVRGGRPSATLAPGTQWSDPVALPTAGGDLVLSLYLPTPGTSTHSA
YQNFVAAGVTCAPDLTPVSTATSWHFLSGVSDVRAGTADSSVVTLDGSEHT
TLADRWPDLLAERLRDGLAGTGVNAGICGNLLLRDPPDPPGSAESAEPAYFGE
SALKRPRDRVQPGAGAVTVLLGVNDLGGPGLAAPSDEVTAEELIACYROLIERAH
EHLGKIYCAITPAGTGTGTYFTPREAVRQVNDWIRISGAFDTVLDFDVLDRPAR
PDHLLPAYDGGDLHPNDAGAAWAFPLDSLR"
complement(3659..4645)
/notes="ORF3 (328 aa)
similar to Y00459-2 Streptomyces griseus regulatory
protein, StrR (350 aa)"
/codon_start=1
/transl_table=11
/product="StrR-like regulatory protein"
/protein_id="BAC76461.1"
/db_xref="GI:30698348"
```

## CDS

```
/translation="MTAALAGNLDQETRSIPISLVPGVSPRLCGDAEHVARLADTD
EALPPIILVERSTLRVIDGHRVLAAKAKGHTTIEVRLFDGAEEAFLLAVRSNMTHGL
```

```
PLSRQDRRAAQRILAQWPHLSDRVAGIAGTAKTVAALRPLAAGTFPQARRGRD
GRRLPDITGRKKAELIAORPOASVREVARHAGISPAVASVRRRLASGRSPVPER
NTQAREAGTGSRAVPGGAAPEPVSVVRPIRPEDPSPLVRLLRDPSLRHKSGRRRL
LRLLQCAVRETBALMAOVTVPCHCTDLVAELAREYADLWAEFAREVARTDG"
complement(4922..5812)
```

## CDS

```
/notes="ORF4 (296 aa), lankacidin biosynthesis protein
similar to AE004625-4 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein B (304 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein B"
/protein_id="BAC76462.1"
/db_xref="GI:30698349"
/transl_table=11
/translation="MILLGTAAAGGPGPQWNCACALCARGRRGLPARSQCVASVGGG
RDWLLNASPDITQLLAPALPFGPDRPTFVRGVLTLDAEVDHGLGLVLRGATGL
TVYAAPVPGALSAELFVRGLLDYAPWRDRTAFGFAAGGLTVTAHPVGTAKPK
YAHAPDPAPWCAYRIEDPATGALVYAPLATWPDGFDLLASATCALLDGTFFSA
GELGTATSSAGASQSLMGHLPVAGPGGSLAALAHRRGLRRIYTHLANTNLLDPSAA
HAAVREAGVEVLDPGSELVL"
complement(5815..6909)
```

## CDS

```
/notes="ORF5 (364 aa), lankacidin biosynthesis protein
similar to AJ277117-5 Gluconobacter oxydans putative pqe
protein (359 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein B"
/protein_id="BAC76463.1"
/db_xref="GI:30698350"
/transl_table=11
/translation="MADPAVGAAGPAGMLIETHRCPLHCPYCSNPDLVIRREAEITCSQ
WTDILTQARELVGVVQHFSGGEPLADPLDPLVGHARRLGAIVNLVTSVGLTAEARH
DLARRGVHQLSLQDADPAAGQAIAGARVHTAKLEAARVTAAGLPLTVNIVLHRGN
IDRTGRVVDLADLDRIELANTQVYWGRLNRALMAPTAAALAAAREAVRHARTY
AGGBELVYAADYVDRPKPCMDGMSQTQTVTPAGDVLPCPAAYATTTLPVENALRR
PLSEIWTASRSFNAVYRTGMRPFCRCRTPHADHGGCRCAQQLTGDAATDPACGL
SPHSLVDAALAEVLDGVPFAFVRGPVPA"
complement(6902..7174)
```

## CDS

```
/notes="ORF6 (90 aa), lankacidin biosynthesis protein
similar to AL603642-200 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein D (98 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein D"
/protein_id="BAC76464.1"
/db_xref="GI:30698351"
/transl_table=11
/translation="MTGLPEPTVRLRPGVRLTRDPARGELALLPERVVVLDATAAV
LAHCGTSLAGIVERLAEEVGEVSAEDVRELLRLAQRVVDLHG"
```

## CDS

```
complement(7171..7899)
/notes="ORF7 (242 aa), lankacidin biosynthesis protein
similar to AE004625-5 Pseudomonas aeruginosa
pyrroloquinoline quinone biosynthesis protein C (250 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein C"
/protein_id="BAC76465.1"
/db_xref="GI:30698352"
/transl_table=11
/translation="MSMSTKEVAAAPSEAEFRORLHALESSYNDRHPHRRMEHGLL
DEGLRLWAANWYQCLPQKDAIVANCPLEVRQWLSRIVHIDGACACAGAEK
WRLAEAVGRDVRDHLVLAQTRFAVDYVDFARRPWLAEASGLTELFSPGGLL
AHRGLRHHYFWIAEEGFEYFTARIIEVVGESRLDLDVARHVAVSREQQACVRAALA
FKKVLNVLSDLYHTGNGATRS"
complement(7926..8030)
```

## CDS

```
/notes="ORF8 (34 aa), lankacidin biosynthesis protein
similar to AL603642-197 Sinorhizobium meliloti putative
pyrroloquinoline quinone biosynthesis protein A (31 aa)"
/codon_start=1
/transl_table=11
/product="pyrroloquinoline quinone biosynthesis protein A"
/protein_id="BAC76466.1"
/db_xref="GI:30698353"
/transl_table=11
/translation="WRTSGKELPAKAWHRPDPFTVIDTGMVTAIFYSR"
8420..10057
```

## CDS

```
/notes="ORF9 (545 aa), possible lankacidin resistance
protein
```

Query Match 17.2%; Score 86.2; DB 1; Length 210614; Best Local Similarity 51.4%; Pred. No. 0.0056; Matches 199; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 112 GCGGGGCGACACCGCGGTCAGTCGTACGTGTACCGGGGAGGTACAGATCCAC 171  
DB 46084 GTGCGGTGGCGGACAGGTCAAGCGGTACCGCGGTGAGCGGGGATCGGGAAGACGAGC 46143

QY 172 TGTCCGCCAGCCCGCGGAACTCTGTCTCTTCGTCATGATCTCGTGGCGGTGTTCCAG 231  
DB 46144 TTGCGCGCCCGCGAGGTAGGCTGCTCGGCTCGAGATCCTCGGCTGATCCAG 46203

QY 232 GCGAAGACAGGCGTATCCACCGCTCGGCGGTGAAGCGGTGCGGGGTCCGACCGGG 291  
DB 46204 GCGAGGACAGCGTGTGTCGGGCGGAACGCTTGGCTCTCTCGGACACGATCGGG 46263

QY 292 ATGTGCTGCGCGGGGTGAGCGCGGCTCTGTGGCGGCGTCTGTGCGACACCGCAGAG 351  
DB 46264 ATGTGCTGCGCGGGGTGAGCGGCGCGCTTCTCTCGCTCACTCGCGATGAGGGC 46323

QY 352 ACCAGTCCGACCGATCGCGGAGAAAGTTGCTGCGGTGGCGTCTTCGCGGTGCGCGG 411  
DB 46324 AGGTGCGGCGGCTGATCGCGAGTACTCGAGGGTCAATTTGCCCTTGGTCGACGCGCG 46383

QY 412 TACGCCACACCGCTTCCCTCGGCTTGAGCGAGTTTACGAGGGGTGAGCGGTGCGGTG 471  
DB 46384 TACCGCGAGGTGAGCAGCGCGGCGCGGGAGTCTGCGAGAACCCCGCGGTGCTGCG 46443

QY 472 CGGATGCCCTCGACCTCGCGCGGCAAC 498  
DB 46444 CGCTGTGTCGCGACCGCGCGGCGGAAC 46470

RESULT 41  
AX535264/c  
LOCUS AX535264 50000 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 1 from Patent WO02068436.  
ACCESSION AX535264  
VERSION AX535264.1 GI:25261906  
KEYWORDS Streptomyces viridochromogenes  
SOURCE Streptomyces viridochromogenes  
ORGANISM Streptomyces viridochromogenes  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Weitauner, G., Muehlenweg, A., Trefzer, A. and Bechthold, A.  
TITLE Avilamycin derivatives  
JOURNAL Patent: WO 02068436-A 1 06-SEP-2002;  
COMBINATORIAL Biopharm AG (DE)  
LOCATION/Qualifiers  
source 1. .50000

Query Match 16.2%; Score 81; DB 6; Length 50000; Best Local Similarity 50.9%; Pred. No. 0.04; Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGTCAGTCTGTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCCGCCAG 181  
DB 15020 CTCGGCCACCGATCACAACTTCCAGGTCGGGAGCGGAAAGACAGGCTGCTCCGG 14961

QY 182 CCGCGGGAACCTCTCTTCGCTGATCTCGCGCGTGTTCACAGGAGAGCA 241  
DB 14960 AGGCCAGGAGTTCGCTTCCCGGCGACCATCGCGTCCGGTAGATCCACGGGAGGACCA 14901

ORIGIN  
Query Match 16.2%; Score 81; DB 6; Length 50000; Best Local Similarity 50.9%; Pred. No. 0.04; Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGTCAGTCTGTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCCGCCAG 181  
DB 15020 CTCGGCCACCGATCACAACTTCCAGGTCGGGAGCGGAAAGACAGGCTGCTCCGG 14961

QY 182 CCGCGGGAACCTCTCTTCGCTGATCTCGCGCGTGTTCACAGGAGAGCA 241  
DB 14960 AGGCCAGGAGTTCGCTTCCCGGCGACCATCGCGTCCGGTAGATCCACGGGAGGACCA 14901

QY 242 GCGCGTAGTCCACCGCGTCCGCGGTGAAACGCTCCGGGTGCGCACCGGATGTGCGTGC 301  
DB 14900 GGAACCTGTGCGGCTCAAGCGGCCGCGCTCTCTCTCGAGACGATCGGGATGTTCGTGC 14841

QY 302 CGGGGTGAGCGCGCTGCTTGGCGGCTGCTGCGCACACCGCAGAGACAGGTCG 361  
DB 14840 CGGCGTGTAGCAGCGAACTTGTCTCGTTCACTCGGCTGCGAGGAGGCTCT 14781

QY 362 GACCGATGCGCAGAAAGTTGCTGCGGTGGCGCTTTCGCCCTCGCGGTACGCCACCA 421  
DB 14780 CGTCCAGGCGCAGTACTGAGATGACGTTGCCCTTCTCGAGCGCGCTAGCCAAAGG 14721

QY 422 CCGCTTGGCTCGGCTTGGAGGTTGCGAGGTCAGAGGCGGAGAGGTCGTCGCGATGCCCT 481  
DB 14720 TGTGAGCCCTTGTCCCGAGGCGGTGAGGAATTCAGGAGTCTGTCACGGGATCTCT 14661

QY 482 CGACGTGCGCGGCAAC 498  
DB 14660 CGTCCGCGCGGCAAC 14644

RESULT 42  
AX535266  
LOCUS AX535266 50000 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 3 from Patent WO02068436.  
ACCESSION AX535266  
VERSION AX535266.1 GI:25261911  
KEYWORDS Streptomyces viridochromogenes  
SOURCE Streptomyces viridochromogenes  
ORGANISM Streptomyces viridochromogenes  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1  
AUTHORS Weitauner, G., Muehlenweg, A., Trefzer, A. and Bechthold, A.  
TITLE Avilamycin derivatives  
JOURNAL Patent: WO 02068436-A 3 06-SEP-2002;  
COMBINATORIAL Biopharm AG (DE)  
LOCATION/Qualifiers  
source 1. .50000

Query Match 16.2%; Score 81; DB 6; Length 50000; Best Local Similarity 50.9%; Pred. No. 0.04; Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGTCAGTCTGTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCCGCCAG 181  
DB 44797 CTCGGCCACCGATCACAACTTCCAGGTCGGGAGCGGAAAGACAGGCTGCTCCGG 44856

QY 182 CCGCGGGAACCTCTCTTCGCTGATCTCGCGCGTGTTCACAGGAGAGCA 241  
DB 44857 AGGCCAGGAGTTCGCTTCCCGGCGACCATCGCTCCCGGTAGATCCACGGGAGGACCA 44916

ORIGIN  
Query Match 16.2%; Score 81; DB 6; Length 50000; Best Local Similarity 50.9%; Pred. No. 0.04; Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGTCAGTCTGTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCCGCCAG 181  
DB 44797 CTCGGCCACCGATCACAACTTCCAGGTCGGGAGCGGAAAGACAGGCTGCTCCGG 44856

QY 182 CCGCGGGAACCTCTCTTCGCTGATCTCGCGCGTGTTCACAGGAGAGCA 241  
DB 44857 AGGCCAGGAGTTCGCTTCCCGGCGACCATCGCTCCCGGTAGATCCACGGGAGGACCA 44916

ORIGIN  
Query Match 16.2%; Score 81; DB 6; Length 50000; Best Local Similarity 50.9%; Pred. No. 0.04; Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGTCAGTCTGTGTCAGTGTACCGGGGAGCTACAGGATCCACTGTCCGCCAG 181  
DB 44917 GGAACCTGTGCGGCTCAAGCGGCCGCGCTCTCTCTCGAGAGATCGGGATGTTCGTGC 44976

QY 302 CGGGGTGAGCGCGCTGCTTGGCGGCTGCTGTCGACACCGCAGGAGACAGGTCG 361  
DB 44977 CGGCGTGTAGCAGCGAACTTGTCTCGTTCACTCGGCGATTCAGAGGAGGAGGCTCT 45036

QY 362 GACCGATGCGCAGAAAGTTGCTGTCAGGTCGCGCTTTCGCCCTCGCGGTACGCCACCA 421  
DB 45037 CGTCCAGGCGCAGTACTGAGATGACGTTGCCCTTCTCGAGGCGCGCTAGCCAGG 45096

QY 422 CCGCTTGGCTCGGCTTCCCGAGGCGGTGAGGAATTCAGAGGCGGAGAGGTCGTCGCGATGCCCT 481





```
CDS
complement(4307..4852)
/gene="aviX5"
/codon_start=1
/transl_table=11
/product="aviX5"
/protein_id="AAK83165.1"
/db_xref="GI:15077438"
/translation="MTSHGINAPCAQGRHYRVPAPDEPLDPAMVPDLTRAPPEDR
AEDRAPWDPDRPALCWLPTSGVMATRRITTTTHIETLVQQLDGDSDASVAAQAQLVD
IGRDAITLIGGLTGLSGFGLTAIEVEFVADPRCPALIALLDADDFSVREWARGP
DRTERPPTSRRSSTTWTFPK"
4931..5794
/gene="aviRb"
4931..5794
/gene="aviRb"
/note="aviRb"
/codon_start=1
/transl_table=11
/product="rRNA methyltransferase"
/protein_id="AAG32066.1"
/db_xref="GI:11141830"
/translation="MAKSRGERTPAARRITSRNARFQQWQALLGNRNKTRAGEFILVM
GVREISLAVHGNVPTLLYDQRELSKWARELLRTVTEQIAMADLLMELGEKNEA
PPEVAVVEMPADLDLPVREDPLVLPDRPTSPGNIGSIIRSDALGAHGLIVAGH
AADYDPKSVRSSTGSLPAPVRVPSPGEVMDWEARRAAGTPIVLGTDHGDGV
FDPTOPTULLIGNETAGLSNAWRILCDYTVSIPWAGSASSLNAANAATAILYEVR
ORISGRATTP"
5809..6159
/gene="aviX6"
5809..6159
/gene="aviX6"
/codon_start=1
/transl_table=11
/product="aviX6"
/protein_id="AAK83166.1"
/db_xref="GI:15077439"
/translation="MYLARASWIRPTISACRSKSPARRPSVAAGSEPSVRASTSPSY
TEDNSAGSGTGREPDASVSIRAVNPSVATSGLTSTGTMTARFISVTCNPNGTGG
WTSTPLQPFRLPR"
6159..6305
/gene="aviX7"
6159..6305
/gene="aviX7"
/codon_start=1
/transl_table=11
/product="aviX7"
/protein_id="AAK83167.1"
/db_xref="GI:15077440"
/translation="MSSKPGQVSPSHRPQTRIRSRACLAASTTSMNGFHRPQLLPM
YCS"
6489..6830
/gene="aviX8"
6489..6830
/gene="aviX8"
/codon_start=1
/transl_table=11
/product="aviX8"
/protein_id="AAK83168.1"
/db_xref="GI:15077441"
/translation="MATPPKWPSEIDYQAPVPGKRAQGFOLLVRIQLPGLHEHKMCS
PAHVLRLHKMPRLVLPKRAADPLTARFTRLHIGGNGLAESRLPMKIPGYEHPG
FRHSLPRTH"
complement(6930..7682)
/gene="aviRa"

Query Match      16.2%; Score 81; DB 1; Length 59816;
Best Local Similarity 50.9%; Pred. No. 0.038;
Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGGTCAGTCGTCTACGCGGGGACGTACAGGATCCACTGTCGCGCAG 181
DB 15020 CTCGCCACCCCGATCACACAATTCACGCGTGGCGAGGAAAGACACGAGGCTGCTCCGG 14961

QY 182 CCGCGCGGAACCTCTGTCTCTTCGCCATGATCTCTGTGCGGTGTGTTCCAGCGGAGAGCA 241
DB 14960 AGGCCAGGAAGTCGCGTTTCCCGGGCGACCATCGCGTCCCGGTAGATCCACGGGAGGACCA 14901

QY 242 GCGCGTAGTCCACCGCGTCGCGGCGGTGAACGCGTCCGGGTGCGCACCGGGATGTGCGTGC 301
DB 14900 GGAACCTGTGCGGCTTAAGCGCCCGGCGCTCTCTCTCGAGACGATCGGATGTGCGTGC 14841

QY 302 CCGCGGGTGAAGCGCGCTCTGTGTGCGCGCGTGTGTGCGCACACCCAGGAGACCAAGGTCCG 361
DB 14840 CCGCGGTGTAGCAGCGCAACTTGTCTCTGTTCACTCGCGATGTCAGGCGAGGACGCTCT 14781

QY 362 GACCGATGCGCGAGAGTTCGTCAAGGTGGCGCTCTTCGCCGTGCGCGCTACGCCACCA 421
DB 14780 CGTCCAGCGCGAGTACTGACAGATGACGTTGCGCTTCTGTGAGCGCGCTGAGCCAGGG 14721

QY 422 CCGCGTTCCTCGGCTTGAGCGAGTTCAGCAGGCGAGAGGTGCGTGTGCGGATGCCCT 481
DB 14720 TGTGAGCGCTTGTGTCGCGAGGCGGTGAGGAATTCAGGAGTCTGTCAAGGACTCTCT 14661

QY 482 CGACGTCGCGCGGAAC 498
DB 14660 CGTCCGCGCGCGGAAC 14644

RESULT 44
AX535308/c 1224 bp DNA linear PAT 22-NOV-2002
LOCUS AX535308
DEFINITION Sequence 45 from Patent WO02068436.
ACCESSION AX535308
VERSION AX535308.1 GI:25261950
KEYWORDS
SOURCE Streptomyces viridochromogenes
ORGANISM Streptomyces viridochromogenes
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Weinbauer,G., Muehlenweg,A., Trefzer,A. and Bechthold,A.
TITLE Avilamycin derivatives
JOURNAL Patent: WO 02068436-A 45 06-SEP-2002;
Combinature Biopharm AG (DE)
FEATURES
Location/Qualifiers
source 1..1224
/organism="Streptomyces viridochromogenes"
/mol_type="unassigned DNA"
/db_xref="taxon:1338"
/note="Teilsequenz avigidna des codierenden Stranges 1.
Das Nucleotid 1 entspricht dem Nucleotid 13784 des
codierenden Stranges 1."

ORIGIN
Query Match      16.2%; Score 80.8; DB 6; Length 1224;
Best Local Similarity 53.1%; Pred. No. 0.12;
Matches 172; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 175 CCGCCAGCGCGGGAACCTCTCTTCGCCATGATCTCGTGGCGGTGTTCCAGGCG 234
DB 1184 CTTCCGAGCGCAGGAAGTCGCTCCCGGGGACACNTCGCTCCGGTAGATCCACGGG 1125

QY 235 AAGAGCAGCGGTAGTCCACCGCGTCGCGCGTGAACGCGTCCGGGTGCGCACCGGATG 294
DB 1124 AGGACAGGAACCTGGTCGGGCTCAAGCGCCCGGGCTCTCTCTCGAGACGATCCGGATG 1065

QY 295 TCGTCCCGGGGTGAGCGCGCTCTGTCGCGCGGCTGTCGCACACCCAGGAGACC 354
DB 1064 TTGTCGCGCGGGGTGAGCAGCGGAACCTTGTCTGTTCACTTCGGCGATGACGGGAGG 1005

QY 355 AGTCCGCGACCGATGCCGCGAGAGTTCGTACGCGTGGCGCTCTTCGCCGTCCGCGGTAC 414
DB 1004 AGCGTCTCTCCAGGCGCGAGTACTGACGATGACGTTGCCCTTCTGTCAGGCGCGGTAG 945

QY 415 GCACACACCGCGTTCGCGCTTCGCGCTTACGAGGATTCAGCAGGCGCGAGGTCGCTGGG 474
```

Db 944 CCAAGGGTGTGAGCCCTTGTCCCGGAGCGGTGAGGAATCCAGGATCGGTACGG 885  
 Qy 475 ATGCCCTCGAGCTCGGCGGCAAC 498  
 Db 884 GACTCCTCGGTCCCGCGGCGCAAC 861

RESULT 45  
 LOCUS AE009716  
 DEFINITION *Brucella melitensis* 16M DNA linear BCT 20-MAR-2003 complete sequence.  
 ACCESSION AE009716 AE008918  
 VERSION AE009716.1 GI:17985017  
 KEYWORDS  
 SOURCE  
 ORGANISM *Brucella melitensis* 16M  
*Brucella melitensis* 16M  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; *Brucella*.  
 1 (bases 1 to 11659)  
 DelVecchio, V.G., Kapural, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltzman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.-J., Haselkorn, R., Kyrpides, N. and Overbeek, R.  
 The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
 Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)  
 11756688  
 2 (bases 1 to 11659)  
 DelVecchio, V.G., Redkar, R.J., Patra, G. and Mujer, C.  
 Direct Submission  
 Submitted (13-NOV-2001) Institute of Molecular Biology and Medicine, University of Scranton, Scranton, PA 18510, USA  
 3 (bases 1 to 11659)  
 Elzer, P.H. and Hagius, S.  
 Direct Submission  
 Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA  
 4 (bases 1 to 11659)  
 Kapural, V., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltzman, E., Selkov, E., Haselkorn, R., Kyrpides, N. and Overbeek, R.  
 Direct Submission  
 Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell Park Drive, IL 60612, USA  
 5 (bases 1 to 11659)  
 Letesson, J.-J.  
 Direct Submission  
 Submitted (13-NOV-2001) Unite de Recherche en Biologie Molculaire, Laboratoire d'Immunologie et de Microbiologie, Université of Namur, 61 rue de Bruxelles, Namur 5000, Belgium  
 6 (bases 1 to 11659)  
 O'Callaghan, D.  
 Direct Submission  
 Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue Kennedy, Nimes 30900, France  
 Location/Qualifiers  
 1..11659  
 /organism="Brucella melitensis 16M"  
 /mol\_type="genomic DNA"  
 /strain="16M"  
 /db\_xref="taxon:224914"  
 /chromosome="II"  
 complement(73..972)  
 /gene="BMEII0820"  
 complement(73..972)  
 /gene="BMEII0820"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ALS OPERON REGULATORY PROTEIN"

gene  
 CDS  
 source

/protein\_id="AAL54062.1"  
 /db\_xref="GI:17985018"  
 /translation="MDIPMELRHLRYFLAVAEENFTRAAKLGIQPPLSQIRDL  
 NEIGVLFHVRPHGAELTAAGAFTLEATATIASSEKAKLAQRAARGESRLSLGFT  
 ASSAFNPISGAIRHPRAKPEVEFLFTMTNNLLIERLVGELDTATIRPLEDPA  
 VMKRLADEPMLVALPSHRLAVYERLPLSALAGEPFLVPRIVGLSLDYDDIAACRM  
 AGFEMNTQEAQPMQSVNLVAANLGSIVPAALQIRLDGVTYRPIEGEPVVARIAL  
 ASMSKLSRFVDNLLISLPRCI"  
 1065..1778  
 /gene="BMEII0821"  
 1065..1778  
 /gene="BMEII0821"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MULTIDRUG RESISTANCE PROTEIN"  
 /protein\_id="AAL54063.1"  
 /db\_xref="GI:17985019"  
 /translation="MSRIEDRSNGIAADNGLTLVEMPEEAKTYLTGMPAYRNASL  
 ALFLSGFATFSLYCVQPLMFADKFSITPTASSLSLSLSTGLAFAIPFAAIAER  
 FQRKLPASLIGSALCTIGCAVVDPMHMLLARSLOGLLGGVPAAVMSYLASEIDP  
 RGLGASGLYIAGNAFGMGAGRVITGTLAEYYSWRPALAVMGLLGLAAAIQFWALLPT  
 SRNPTFGMEPATYRPG"  
 1643..2365  
 /gene="BMEII0822"  
 1643..2365  
 /gene="BMEII0822"  
 /codon\_start=1  
 /transl\_table=11  
 /product="MULTIDRUG RESISTANCE PROTEIN"  
 /protein\_id="AAL54064.1"  
 /db\_xref="GI:17985020"  
 /translation="MAPSPCRHGIARSGSSNRLGIVAYVQKFAQNRMGARFHVSAW  
 LSHLRNPAFLMGLIAGFTVMGPFVITYNIGFRLIEPPNLSQTELAGFIPFVYLFGLA  
 ASWAGRIGDRGLGHTMPLIALAYBAAGCLITLSASLPYIVGGIFLTVGDFMGSVA  
 SALVRLALBTKHASSLYLLAYVLGSSIAAGSYFWTADGNWAVADFTFAMLALAP  
 AALAAAFABEGQTDKPLTF"  
 2492..2659  
 /gene="BMEII0823"  
 2492..2659  
 /gene="BMEII0823"  
 /EC\_number="2.7.1.30"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GLYCEROL KINASE"  
 /protein\_id="AAL54065.1"  
 /db\_xref="GI:17985021"  
 /translation="MSGYILAIDQGTSTRMLFDRNRRVVGLOQEQEFTQHPSSGWW  
 BDAEEIWKSV"  
 2918..3988  
 /gene="BMEII0824"  
 2918..3988  
 /gene="BMEII0824"  
 /EC\_number="2.7.1.30"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GLYCEROL KINASE"  
 /protein\_id="AAL54066.1"  
 /db\_xref="GI:17985022"  
 /translation="MLNRPGLRRRAQKQVCFQCTIDSWLIYKLTGCAKAVHTDATNAS  
 RTLIYHIGENRWDDLLDIGIPAAFLPEVKCAADFGMTDFPALFGVSIPTLVGAGDQ  
 QAIVGNACPEPGMKSTYTGCCFALNTGTRVTSNRLITTYAILDGVTTVALEG  
 SIVTAGAVALWRDEMGMFISVASEVSALAEKADPNQRIYLPVPTGLGAPYWDAAEG  
 AIFGLRTGTGRAAFARALAEVAYOTFDLEAMQGDWKGATNHTVLRVDGGMVADMT  
 MQLADILINAPVDRPVLETTVLGAAMLAASRAGIWPDRKGFSEKQWQDCRFEPAMPE  
 KERESATAGWEDSVSRCLTRPQ"  
 4105..5385  
 /gene="BMEII0825"  
 4105..5385  
 /gene="BMEII0825"  
 /codon\_start=1  
 /transl\_table=11  
 /product="URACIL PERMEASE"  
 /protein\_id="AAL54067.1"

/db\_xref="GI:17985023"  
 /translation="MGYPSPNRLTSNAVLVDELPAALAAVPMGIQHLAMSGSTIVA  
 PLLMGFDNVAFFSGITLFLFLITAGRPVSLGSSPAFLAVIATGYSQSGPPI  
 IALAGLGIACGLYSLIGLIVMELGTGWKELMPAVTGAIGVSGINLAPVAGL  
 KTGHTSIALTLWLCMAISAYGRTRLSVLFPALLFGILLVAGNGLIGVPGDI  
 FTAVKNAWGLPHTTFAFNQAVTIIAPVAVIIVAEHGHKALGSIIGRNDY  
 FGAILQTPAVLAGLAVSFLGVIASAMARWIIVKVDFAISRNLFTVGVALLIFGAGD  
 FTILGNFALGCIATSTLAALVQLLIGRSDS"  
 complement (5581..6237)  
 /gene="BMEII0826"  
 complement (5581..6237)  
 /gene="BMEII0826"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Hypothetical Cytosolic Protein"  
 /protein\_id="AAL54068.1"  
 /db\_xref="GI:17985024"  
 /translation="WIDLCALANPGRPLELCLGASHDDIEIGCGTLLHITGTPL  
 HVENCVLGNQRRARASARDFELGRMDRQLHVAFADSIFPAESRAIKEMLIAGR  
 GRQPDIVFTYHGDAHQDRTVNELTWNLPDQILLEYEIPKWDGLGRPNAYSLP  
 ASIWKRIDLLKMGHGTORSQDFGPFTRGLARLGMCEAPETAYAEAFHAKRLIF  
 "  
 complement (6234..7121)  
 /gene="BMEII0827"  
 complement (6234..7121)  
 /gene="BMEII0827"  
 /EC\_number="2.7.7.33"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GLUCOSE-1-PHOSPHATE CYTIDYLTRANSFERASE"  
 /protein\_id="AAL54069.1"  
 /db\_xref="GI:17985025"  
 /translation="MGRHSHSPCHLDRSGTRYGGALINKVVLFCGGIGTRIREYS  
 ENIPKMLPLGHPQLHVRMEYISDYGHNDVFLCIGYKANVKKDFITIPQIPADCV  
 VSDGNVQLLEQDRLVDWSTVLTDTWRNTGRLSARSHWNEDFLANYSDGLSD  
 VLQDNTQFRASGVACFLAVRPPLTYHLADIAGGDSVAFRTSNTSDIWIINGYFL  
 FLKEIFDYMRGSEILVFPFSLRLAENQLMAYKYDGFWRSDTLRDWQTLDMVVEKD  
 MPWKXKAEQAANAIAL"  
 complement (7046..7612)  
 /gene="BMEII0828"  
 complement (7046..7612)  
 /gene="BMEII0828"  
 /EC\_number="2.1.1.-"  
 /codon\_start=1  
 /transl\_table=11  
 /product="POSSIBLE S-ADENOSYLMETHIONINE-DEPENDENT  
 METHYLTRANSFERASE"  
 /protein\_id="AAL54070.1"  
 /db\_xref="GI:17985026"  
 /translation="MAQHHLKVIDVEELPTHGGLRVLVLAHGSKRKRGVPRVASLLK  
 REESFLNEISTYEQFAEKTRTKRDLISFLIAAKNAGKRICGYGAPKGNILNYCG  
 IGTDPFGTVDRNPYKHGRFTPGMHIPIYDVSAIDNRPDYVILIPWFKDEILRQW  
 HVVEGAKFIPIPHVTLIDPALVTEER"  
 complement (7684..8346)  
 /gene="BMEII0829"  
 complement (7684..8346)  
 /gene="BMEII0829"  
 /EC\_number="2.1.1.-"  
 /codon\_start=1  
 /transl\_table=11  
 /product="POSSIBLE S-ADENOSYLMETHIONINE-DEPENDENT  
 METHYLTRANSFERASE"

Query Match 15.6%; Score 78; DB 1; Length 11659;  
 Best Local Similarity 51.1%; Pred. NO. 0.15;  
 Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 140 TCACGTGTACGGCGGAGGTACAGATCCACTGTCGGCCAGCCCGCGGGAACCTCTGCT 199  
 DB 7078 TCAGGTGACATGGGAATGGGAATGATGAATTCGCCGCCCATTCACAGCATGCTGCA 7137  
 QY 200 CCTTCGCCATGATCTCGTGGCGTGTTCACGGCGAAGACGCGGTAGTCCACCGCGT 259

Db 7138 TCTGCCGGAATATCTCGTCTTGAATTCATCCAGGATGAGATGATGTCGGCCGT 7197  
 QY 260 CGGCGGTGAACGGTCCGGGGTCCGACCGGATGTGCGGGGGTGAACCGGCCCT 319  
 Db 7198 AATTGTGATCGCCGAAACATCGTAATCGAATGTGCAATTCAGCGCGTAAGCGCCCT 7257  
 QY 320 GCTTGGCCGGCGTGTGTGCGCACACCCAGGACACAGGTCCCGGACCGATGCGCGCAAGCT 379  
 Db 7258 GCTTGTAGGATTCGATCCACGGTGAACCCAGAAATCCGTGCGGATCGCGCAATAGT 7317  
 QY 380 TGCTACGGTGGCGCTCTTCGGCGTCCGCGCTAGCCACACCGCTTGCCTCGGCCCT 439  
 Db 7318 TCAAAACGGTGTTCGCCCTTTCGCCGGCGCCATAGCCACAGATGCGTTCGCCGCAATTT 7377  
 QY 440 TGAGCGAGTTTCAGCGGGCGAGCAGGTGCGTGGATGCCCTCGACGTCGGCGCGCAA 497  
 Db 7378 TGGCGCAATCAGGAGGACAGAGTACAGTTCGTTGGTGGCGCGCTCTTTTCGGCAA 7435

RESULT 46  
 AE014542/c  
 LOCUS 11852 bp DNA linear BCT 02-OCT-2002  
 DEFINITION Brucella suis 1330 chromosome II section 38 of 108 of the complete  
 genome  
 ACCESSION AE014542  
 VERSION AE014542.1 GI:23463781  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Brucella suis 1330  
 Brucella suis 1330  
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 Brucellaceae; Brucella.  
 REFERENCE 1 (bases 1 to 11852)  
 Paulsen, I., Seehadi, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F.,  
 Read, T.D., Dodson, R.J., Umayam, L.A., Brinkac, L.M., Beanan, M.J.,  
 Daugherty, S.C., Deboy, R., Durkin, A.S., Kolonay, J.F., Madupu, R.,  
 Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J.A., Van  
 Aken, S.E., Riedmiller, S., Tettelin, H., Gill, S., White, O.,  
 Salzberg, S.L., Hoover, D.L., Lindler, L., Halling, S.M., Boyle, S.M.  
 and Fraser, C.M.  
 The Brucella suis genome reveals fundamental similarities between  
 animal and plant pathogens and symbionts  
 Proc. Natl. Acad. Sci. U.S.A. 99 (20), 13148-13153 (2002)  
 12271122  
 REFERENCE 2 (bases 1 to 11852)  
 Paulsen, I., Seehadi, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F.,  
 Read, T.D., Dodson, R.J., Umayam, L.A., Brinkac, L.M., Beanan, M.J.,  
 Daugherty, S.C., Deboy, R., Durkin, A.S., Kolonay, J.F., Madupu, R.,  
 Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J.A., Van  
 Aken, S.E., Riedmiller, S., Tettelin, H., Gill, S., White, O.,  
 Salzberg, S.L., Hoover, D.L., Lindler, L., Halling, S.M., Boyle, S.M.  
 and Fraser, C.M.  
 Direct Submission  
 Submitted (14-AUG-2002) The Institute for Genomic Research, 9712  
 Medical Center Dr., Rockville, MD 20850, USA  
 FEATURES  
 source  
 1. 11852  
 Location/Qualifiers  
 /organism="Brucella suis 1330"  
 /mol\_type="genomic DNA"  
 /strain="1330"  
 /db\_xref="taxon:204722"  
 /chromosome="II"  
 89..1468  
 /locus\_tag="BRA0432"  
 89..1468  
 /locus\_tag="BRA0432"  
 /note="similar to Gp:15158361; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="glycosyl transferase, group 1 family protein"  
 /protein\_id="AAN3362.1"  
 /db\_xref="GI:23463782"

```

/translation="MRGQAIINWQYISCPRYTCTRLRGSPFSPEVRPEPSRKGPGV
QKSLAMKAVRPHIGTYSVFTLRSLASLRSQWDLFWLAAGAPGEAVAVPREQGL
EDMRGGDIIDRMVGLDEETRAKRVDFLRENFRAVFNVLVSHRFTNLYVLPADI
LHVIVHSITPTGTYAAQAIRDRHVATGVSKRCDLVRGVGDFPSRTLVPNGLTP
DRVRCAPAEIPEGGRLLRLYLGRIDDSKGVFWLPAILRLQCDYHLTVAGDGPDLAA
LRIKAPGNRISFTGCVAPSDVSVLQSHDVMVMPSEYEGFSGTLIAMSQGCPPVA
SLIPGVTDTIVTDGEGRLFPPTIGNCQQAQOIERLAADPRLRVANGAGURKVIAREFD
SLVGEAVAGLLQHLRENRPITPPLAQSOMSPNALHKNLSRLPQPVKNWLRVRE
RHTVWSTA"
gene
1506. .2867
/locus_tag="BRA0433"
CDS
1506. .2867
/locus_tag="BRA0433"
/notes="similar to GP:15158369; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamate-1-semialdehyde-2,1-aminomutase,
putative"
/protein_id="AAN33627.1"
/db_xref="GI:23463783"
/translation="MKLFTSQFNSQLLAKAHALPGCHTYAKGDDQYVYLAPGFI
QRSGNHVDVGVHEYIEYMGNGRAVGLGHAYPPVAVREALQDCNFPSPSAELD
CAESFLELEGEMVKFCCKGSDATSGALRLARAYVVRVRELVACCDHPFSTDDWFIG
TKMSAGIPASVALTATPRYNDIASVEALFEDYPCRIAAIMEPARADEPQDNFLHE
AMRIAHENGALFIDEMITGFRWHRGAQLYGIYVDLSCFCGKALNGFAISALAKGA
EYVRLGGLHTDPFRVFLSTTTHGAEATHAMAAAIATMAIYRDEPVIERLYEQGAKLAA
GNEALAGRLQDHVRLSCRPCCLTYSTLDGNGMPSQAFRILFLOETIRRGVIMPSLV
VYTHSDADIARTIDAVHGALGIYVRALNDGVESCLNGRPSQVYRNEFPASSCPP
RL"
gene
2846. .4180
/locus_tag="BRA0434"
CDS
2846. .4180
/locus_tag="BRA0434"
/notes="similar to GP:15158360, and GP:15158360; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN33628.1"
/db_xref="GI:23463784"
/translation="MPTAKIVRGIALFLFLLAPMTGGARAIADTOPFLQVSGDRHRLQ
DAAETCLLGGANSLIAELRDEREAVLQDRRRGNFNAVILVLEHQFARNAPANT
YGRPFKARAFALNPQYFDHAHWIGRAELGIVFLASATYLGANGGGQYPMWEM
AGPKRWAGEVIARFAGYRNIIWVLGGDFDAPKRLVSLAEGIAVLDAIQTVH
SERDITABHAGALQALDVTYDDVHKAVLARTAGAMEVILLETLYNEHGTDA
CTRIRNAYGALLAGAGQFQFGSPITWHPSEAPGLFDGNGWRKAPDPSGARSNGVLQNL
FSRLDWTLOPDRKKITDRPGFYAARLPETGLAVTYGNGFALRGKAVDAGDAALWF
DPASGKTIIVAGAPARENDLLVHPFASRNADGSHDWLLIGRMERLAFLOKE"
gene
4301. .5362
/locus_tag="BRA0435"
CDS
4301. .5362
/locus_tag="BRA0435"
/notes="similar to GP:15158370; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="epimerase/dehydratase family protein, putative"
/protein_id="AAN33629.1"
/db_xref="GI:23463785"
/translation="MKVLVTGSGYIGSVYVWVPLMDAGHESVGYDIGLYHCLFREGG
TVMNVPTIKRDVDSYPRDLGFDVHILHAAISNDPLNGLNEDLVYAHNHSVAMAK
AAKLAGVGFPLASCSNYGISDAELIDETGLKPVYAYGRSKVRAEODEIKLADASF
CPVYLRPATYGVSPLEFDIYNNLVAVAWTKGLIYLSKDGTPRPVPIHIDISRAF
MAAMEAPREKVNREAPNVGSTAHYNRIDRIAGVADNPDCRIEYVAQAGDTSIRV
SEKLAHLVPEAKPQMDAVAGTRELIAAYSRTLTVEEFEGPRFORIAHRKLIAQGV
LDADRMDAHSPLPRAVA"
gene
5362. .6714
/locus_tag="BRA0436"
CDS
5362. .6714
/locus_tag="BRA0436"
/notes="similar to GP:15158371, and GP:15158371; identified
by sequence similarity; putative"

```

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAN33630.1"
/db_xref="GI:23463786"
/translation="MDSGAPGRAGGARAPODGAHIEHLAARLYPICRSLAGDGVYR
ETALIGEYPLIHIEHVEPTGTPLYDMKAPQEWIIRACIADLAGRIYDFAHNHLVY
NFSMPVARNSLDALKNHIIHTLDPQDPLIFIRYGESMGFCMAHALLAMGDEGFD
VWDAERDGSGLVGFPIHRGQDTADLLSAHLCHPSLANDSCGLGGGPNYKRRGNAE
RRTRLTYRLFGPATFALAWLNEDNLDHVRHGLVLSCLGGGPNYKRRGNAE
IDRIMDHVLAHCLGSLGAAMHDFWYQDERQFQSPGNLPGMFQSRGLYGRFPYHTS
ADNLDFIKPEYLSQSNMVMQTTIEIARNVWPLNMGPKBPOLGRRLYSGMGDSRA
AQNAMAMLVNLADGYSILDMAERSKLFPAMLADAADRLDAGLIKETASQDLPAK"
gene
complement(6677. .7201)
/locus_tag="BRA0437"
CDS
complement(6677. .7201)
/locus_tag="BRA0437"
/notes="similar to SP:29202, and PID:148782; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dtdp-4-dehydrorhamnose 3,5-epimerase"
/protein_id="AAN33631.1"
/db_xref="GI:23463787"
/translation="MHFTPLNINAMSIERILKDRSGEAFYVYCEAAFAERKLETHF
PQHSLSRSRKGILRGLHYNQNEHSETKLVTCIQIYWDVLDLRPHSPYRWAGYE
LSAENGVLKIPGSCAHGFQTLDDVLIRINISKPYADPHAAGRIYDDPADDIPWPEK
PSVISQKDLGWPF"
gene
7522. .8823
/locus_tag="BRA0438"
CDS
7522. .8823
/locus_tag="BRA0438"
/notes="similar to GP:15158375; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="methyltransferase, putative"
/protein_id="AAN33632.1"
/db_xref="GI:23463788"
/translation="MTIQAADISRDSHTVQARTGCRLCGKPLSHSFVDLCWSPCESF
VTARLEDMHEPYTPHQAQCDHCFVLQKEYFSPADIPTYEYAFSSFATSWVSHAKNY
CDEITERLEDNDSFVVEIASNDGYLLQHPFKNIPILGIEPAANVARTAIKGVPL
VDFPGSLALEMIGERKADLIIGNVLAQVPLNDPVRGMQLLKPEGVITLTFPHL
VNIIEQOFTIYHEHFSYFSLTIRFMAQHHLKVIDVEELFTHGSLRVYLAHGS
KRKVGPRVALLKREESFGLNEISTYEOPAEKTRTKRDLLSFLIAAKNAGKRICYGV
APCKGNTLANYCGITDGLFGTVDNRNYPYKHGRTPGWHIPIYDVSADNVRPDIIL
PWPKDBIIRQMQHVVWENGAKFIPHPVTLLDIPALVTEER"
gene
8926. .9635
/locus_tag="BRA0439"
CDS
8926. .9635
/locus_tag="BRA0439"
/notes="similar to GP:15158376; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="nucleotidyltransferase family protein"
/protein_id="AAN33633.1"
/db_xref="GI:23463789"
/translation="MKVVLFCGGIGTRIREYSNI PKPMIPLGHQPIHLHVMYYSYD
GHNDFVLCYKANVQDFLTIRPOTFADVCVSDGGRNVOLLBEVDRDWSVTLDTG
IWNICERLASBSHWNNEDI FLANYSDGLSDVLDLDMTORFRASGVACFLAVRPL
TYHLDAIEGCDVRAFRTNTSDIWINNGYFLFKELFDYMRSGEELVLEPSELLAE
NQLMAYIDGDFWRKSMOTLRDWMQTLDEDMVEKGMFWKKHKAQRAANIAL"
gene
9632. .10288
/locus_tag="BRA0440"
CDS
9632. .10288
/locus_tag="BRA0440"
/notes="similar to GP:15158377, and GP:6066675; identified
by sequence similarity; putative"
/codon_start=1

```



```

/ gene="bme16"
/ note="putative uracil permease"
/ codon_start=1
/ transl_table=11
/ product="Bme16"
/ protein_id="AAL27669.1"
/ db_xref="GI:16740538"
/ translation="MGYFSPRLTNSAVLPDRERPAAPVPMGIQHLIAMSSTIVA
PILMGDPNVAVFFSGIGILLFLITAGRPVPSYLGSSFAIVVIAATGYSGSGNPN
IAIALGGIACGLVSLIGLIVNFGTGWKELMPPAVTGAIGVSLGNLAPVAIGQI
KGTGATSLALTVLCMAMISAYGPGTRRLSVLPAVLFYGLLVLIAGNLGIVFQID
FTRVAAWAFGPHFTPAFPAQVAVLIAVAIVLVAENLGHKALGISTIGENMDGSPK
GRGFFDGLTMSISAGGVTVAENIGVMAITKVSTLIFVIAAMALITLIGMSPK
FTALINFGALGIGATSLAALVLYQLGIGRSDS"
/ complement(3675..4331)
/ gene="bme17"
/ complement(3675..4331)
/ gene="bme17"
/ codon_start=1
/ transl_table=11
/ product="Bme17"
/ protein_id="AAL27668.1"
/ db_xref="GI:16740534"
/ translation="MIDLCALANPGRLPBLCLGASHDDIEICGGTLLHLETGTL
HVECVLSNGEQRRAEASARDLFGMDRFLHVASFADSYFPAESRAIKEWLIAQR
GRQPDIVTHYGHDAHQDRHTVNLNWLFDQLILEYIPIKWDGLGRPNVYALP
ASIMERKIDLLAKHFGTQRKDFGTFETFRGLARLRGMECRAPETVAAEFARKLRF"
/ complement(4328..5137)
/ gene="bme20"
/ complement(4328..5137)
/ gene="bme20"
/ codon_start=1
/ transl_table=11
/ product="Bme20"
/ protein_id="AAL27669.1"
/ db_xref="GI:16740535"
/ translation="MKVLFVFCGLGTRIREYSENIPKPMIPILGHQFPLHHVMEYSDY
GNDFVLCIGYKANVVDFFLTIRPQTADCVSDGSRNVQLLEVDWDSVTLIDTG
IWRNIGERLWSARSHVMDIEDIFLANSGLSDGLVDLDMTQFRASGKACFLAVRPL
TVHLADIAGDGVRAPTNTSDIINGGVFLFKEIPDYMEEGELVLEPESLIAE
NQLMAKYDGEWSDTLDMQTLMDVWKGDPMPKHKAEQRAANIAL"
/ complement(5140..5441)
/ gene="bme21"
/ complement(5140..5441)
/ gene="bme21"
/ note="putative C-methyl transferase"
/ codon_start=1
/ transl_table=11
/ product="Bme21"
/ protein_id="AAL27670.1"
/ db_xref="GI:16740536"
/ translation="MTIQADISRDHSTVQARTGCRCLKLRHSFVLDIGMSPPCSP
VTAERIDLENDPYPVQCHDFVQLVKEPADICEYAYFSSFATSWYSHAKNY
CDEITRLENDPNSFVVEIASNDGYLLOHFLPKNIPILGIEPAANVARTALAKVPTL
VDFPFGSLAEMIGEGKADLIGNNVLAQVPLDVFYGMQLLLKPGVITLLEPHL
VNLIEQNDFTIYEHFSPSLITIRFMAQHHLKVIDVNEELFTGGSIRVTLAHGS
KRVGPRVSLKREESPGNEISTEQFAEKTRRTKRLDLSFLIAAKNAGKRICGY
APKGHTNLNYCGIGDFLGTVDNPNPKHGTTPGMHPIYDVSAINDNRPDYLLIL
PWNFKDEIRQMQHVVEGAKFIIPHPVTLIDPALVTEER"
/ gene="bme22"
/ complement(6731..7285)
/ gene="bme22"
/ note="putative epimerase"
/ codon_start=1
/ transl_table=11
/ product="Bme22"
/ protein_id="AAL27671.1"
/ db_xref="GI:16740537"
/ translation="MQKCSNRGQAHFSAHQHQRVHVRAGTAKDBERGWFARVYCEAA

```

```

FAERKLETHFPQHSLSRBEKGTILRGHLHYQNEPHSETKLVTCLOGIWDVLDLRPHS
PTVRWAGVLSAENGVLQYIPBCGACGFTLTDVLRIMISKYPADPAAAGIRYDD
PALDIPWPEKPSVISQDLGWPL"
/ complement(7248..8561)
/ gene="bme23"
/ complement(7248..8561)
/ gene="bme23"
/ codon_start=1
/ transl_table=11
/ product="Bme23"
/ protein_id="AAL27672.1"
/ db_xref="GI:16740538"
/ translation="MAAPQDPAIHLARLVPICRSLAGDVGRETLLAIGEYLPILH
IHEVPTGTFLYDKNAPOEWLIRACIADLAGRIVDPARNLHVPMVPMVRAEMSLD
ALKNHITLFDQDPLIPIRYCYHGESWGFCAHNAALLAMEDGFEDVVIDAERDGSIV
FGFIRHGQTDATFLLSAHCPISLANDNCGLALLALLGEAKSRKRLTYRLLPFGP"
Query Match 15.6%; Score 78; DB 1; Length 36502;
Best Local Similarity 51.1%; Pred. No. 0.11;
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;
Qy 140 TCACGTGTACGCGGGGACGTACAGGATCCACTGTCCGCCAGCCCGCGGAACTCTGCT 199
Db 5172 TCAAGGTGACATGGGGAATGGGAATGATGAATTTCCGCGCCCATTCACGACATGCTGA 5231
Qy 200 CTTTCGCCATGATCTCGTCGCGGTGTTCCAGGCGAAGACAGCGCGTAGTCCACCGCGT 259
Db 5232 TCTCGCGATAATCTCGTCTTGAATTCATGCGCAGATGAGAATGCTAGTCCGCGCGGT 5291
Qy 260 CGGCGGTGAACCGTCCGGGTGCGCACCGGATGCTGTCGCGGGGTAGCCGCGCT 319
Db 5292 AATTGTGATCGCCGAAACATCGTAAATCGAATGTGATTCACAGGCGTAAAGCGCGGT 5351
Qy 320 GTTTGGCGGGTGTGTCGCACACCCAGGAGACAGGTCCGACCGATGCGCAGAAAGT 379
Db 5352 GTTTGAGGATTCGATCCACGGTGAACCCAGAAATCCGTGCGCATGCCCAATAGT 5411
Qy 380 TCGTACCGTGGCGCTCTTCGCGCGGTGACGACACCGCTTTCGCTTCGCGCT 439
Db 5412 TCAAAAGCGTGTTCCTCCCTTCGCGGGCGGCATAGCCACAGATGCGTGTTCGCGCATTT 5471
Qy 440 TGACGAGTTTCAGCAGCGCGAGCGGTGCGTGGGATGCCCTGACCTCGCGCGCGAA 497
Db 5472 TGGCGGCATCAGGAAGACAGAAAGATCAGCTTGTGTCGCGCGCTTTTCGCGCAA 5529
RESULT 48
AF329398/c 42291 bp DNA linear BCT 13-DEC-2002
LOCUS Streptomyces roseochromogenes subsp. oscitans clorobiocin
DEFINITION biosynthetic gene cluster, complete sequence.
ACCESSION AF329398
VERSION AF329398.1 GI:24940583
KEYWORDS Streptomyces roseochromogenes subsp. oscitans
SOURCE Streptomyces roseochromogenes subsp. oscitans
ORGANISM Streptomyces roseochromogenes subsp. oscitans
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (Bases 1 to 42291)
AUTHORS Pojer,F., Li,S.M. and Heide,L.
TITLE Molecular cloning and sequence analysis of the clorobiocin
biosynthetic gene cluster: new insights into the biosynthesis of
aminocoumarin antibiotics
Microbiology 148 (Pt 12), 3901-3911 (2002)
JOURNAL 12480894
PUBMED 2 (Bases 1 to 42291)
AUTHORS Pojer,F., Li,S. and Heide,L.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-2000) Pharmazeutische Biologie, Pharmazeutisches
Institut, Auf der Morgenstelle 8, Tuebingen 72076, Germany
FEATURES
source
1..42291
/organism="Streptomyces roseochromogenes subsp. oscitans"

```





/transl\_table=11  
/product="clog"  
/protein\_id="AA65221.1"  
/db\_xref="GI:24940594"  
/translation="MHTAVIIGTGMIGTSGISLRSKQGVDTYLMDSVPAIRISEAVG  
AGWAEPEAVDLAVVAVPAQVAVIASHQSRVAFYVVDVAGVKVSLRELBALGC  
DLATVVGHPPLVGRPGSPPLAADDLPGRPWALVPAATDASALNCALVLCGAI  
PVLDAEADRAILGTLVPOVLTAAARLTADSIARLLGLGWWSRIPOLIGVDS  
TSWTQALANAPVYSELEKSLDLITLLETLRAVDGSGSPAEPPDGHLLFIORGID  
GNNRIPRGITGTEGLADIDVFEDRPELTARLPDRVASGVNMGINISQRDPDLD  
RVTTSVTPRKAEDLLHLRERKWPADA"  
/gene="clog"  
11746..12705  
11746..12705  
/gene="clog"  
/codon\_start=1  
/transl\_table=11  
/product="regulatory protein clog"  
/protein\_id="AA65222.1"  
/db\_xref="GI:24940595"  
/translation="MSNSGDEBITSIPIKTKRVRVSIISLLPPSKLRSGETEHI  
RALAEETLPIVVRHETRRVIDGMHRLRAAFRGDLEIVFVDSADVFVLAVE  
LNRAGLEPITMAERKSAARINDSPHSDRLKARTTGLASTIASLRSSISGTVR  
RTGQDRIRPDGTGRQASDLAQNPNASREVAKTARISVGFASDVRALRRGEP  
AVTARQAVKLKRPAAATQRTSPDYRVLSLRKDPISLRTDVGRRLLDLGAPGS  
VQIAQIADISVPEHCRIVVVDWARECAAAWQLADLAGRDTA"  
/gene="clog"  
12920..13135  
12920..13135

Query Match 15.6%; Score 78; DB 1; Length 42291;  
Best Local Similarity 51.1%; Pred. No. 0.1;  
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;  
QY 140 TCACGTGTACCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAACCTCTGCT 199  
DB 37841 TCACGGTGTTCGGGTATGGGACATATAACGGCGCGCGTTTCCAGAACGACTTCT 37782  
QY 200 CTTTCCGCGATGATCTCGTCGGCGTGTTCAGGGGAGACAGGGGTATGTCACCGCGT 259  
DB 37781 CTTTGTGTATGATCTCCGGCGCGTAGTTCAGGCGCAACAGAGTGTAGTTCGCT 37722  
QY 260 CGGGCGTGAACCGGTCCGGGTCCGACACCGGAGTGTGCTGCGGGGGTGAGCCGCGCT 319  
DB 37721 GGTGCGTGTCTTCGGTGACAGATCGAATCTGGACCTGGTAGATTTTCCCT 37662  
QY 320 GCTTGGCGCGGTCTGTCTGCACACCCAGGACCGAGTCCGACCGATGCCGCGAAGT 379  
DB 37661 GTTTCAGAGTGTGTGTCTCACTACAGAACCGGATCTCTGATGCCCGCGCATCGG 37602  
QY 380 TCGTCACGGTGGCGCTCTTCGGCGTCCGCGGTACGCCACACCGCTTGCCTTCGCGCT 439  
DB 37601 CGAGCATGTGACTCCCTTTAGTTGGTGGCGCGGTATCCAGCACATGATTTGCCCTTCCCA 37542  
QY 440 TGACGCGAGTTTCAGCAGGCGGAGCGAGTCCGTGCGGATGCCCTCGACCTCGGCGCGAA 497  
DB 37541 CCACGCGCATACCACTTCAGTTCGGATCTGTGCGGAACTCGTTCGTACGCTCGCGGAA 37484

RESULT 49  
AX707118/c 42291 bp DNA linear PAT 04-APR-2003  
LOCUS  
DEFINITION Sequence 4 from Patent WO03014352.  
ACCESSION AX707118  
VERSION AX707118.1 GI:29563423  
KEYWORDS  
SOURCE Streptomyces roseochromogenes  
ORGANISM Streptomyces roseochromogenes  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Heide, L. and Li, S.M.  
TITLE Nucleic acids for aminocoumarin biosynthesis

JOURNAL Patent: WO 03014352-A 4 20-FEB-2003;  
Universitaet Tuebingen (DE)  
FEATURES Location/Qualifiers  
source 1. 42291  
/organism="Streptomyces roseochromogenes"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:67357"  
ORIGIN  
Query Match 15.6%; Score 78; DB 6; Length 42291;  
Best Local Similarity 51.1%; Pred. No. 0.1;  
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;  
QY 140 TCACGTGTACCGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAACCTCTGCT 199  
DB 37841 TCACGGTGTTCGGGTATGGGACATATAACGGCGCGCGTTTCCAGAACGACTTCT 37782  
QY 200 CTTTCCGCGATGATCTCGTCGGCGTGTTCAGGGGAGACAGGGGTATGTCACCGCGT 259  
DB 37781 CTTTGTGTATGATCTCCGGCGCGTAGTTCAGGCGCAACAGAGTGTAGTTCGCT 37722  
QY 260 CGGGCGTGAACCGGTCCGGGTCCGACACCGGAGTGTGCTGCGGGGGTGAGCCGCGCT 319  
DB 37721 GGTGCGTGTCTTCGGTGACAGATCGAATCTGGACCTGGTAGATTTTCCCT 37662  
QY 320 GCTTGGCGCGGTCTGTCTGCACACCCAGGACCGAGTCCGACCGATGCCGCGAAGT 379  
DB 37661 GTTTCAGAGTGTGTGTCTCACTACAGAACCGGATCTCTGATGCCCGCGCATCGG 37602  
QY 380 TCGTCACGGTGGCGCTCTTCGGCGTCCGCGGTACGCCACACCGCTTGCCTTCGCGCT 439  
DB 37601 CGAGCATGTGACTCCCTTTAGTTGGTGGCGCGGTATCCAGCACATGATTTGCCCTTCCCA 37542  
QY 440 TGACGCGAGTTTCAGCAGGCGGAGCGAGTCCGTGCGGATGCCCTCGACCTCGGCGCGAA 497  
DB 37541 CCACGCGCATACCACTTCAGTTCGGATCTGTGCGGAACTCGTTCGTACGCTCGCGGAA 37484

RESULT 50  
SCO939126 295150 bp DNA linear BCT 11-FEB-2003  
LOCUS  
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 23/29.  
ACCESSION AL039126 AL009199 AL023517 AL023702 AL034414 AL034446 AL034492  
AL035212 AL035478 AL049727 AL132824 AL132997 AL1599708 AL1591522  
AL645882  
AL939126.1 GI:24413894  
KEYWORDS Streptomyces coelicolor A3(2)  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE 1  
AUTHORS Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L.,  
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,  
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,  
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,  
Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,  
Rabinowitsch, E., Rajandream, M.A., Rutherford, K., Rutter, S.,  
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,  
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrall, B.G.,  
Parkhill, J. and Hopwood, D.A.  
TITLE Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)  
JOURNAL Nature 417 (6885), 141-147 (2002)  
MEDLINE 21996410  
PubMed 1200953  
REFERENCE 2 (bases 1 to 295150)  
AUTHORS Bentley, S.D.  
DIRECT SUBMISSION  
TITLE Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces  
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
JOURNAL On or before Oct 26, 2002 this sequence version replaced  
COMMENT

```

gi:20520744, gi:20520662, gi:20520663, gi:20520667, gi:20520669,
gi:20520761, gi:20520816, gi:20520670, gi:20520820, gi:20520846,
gi:20520823, gi:20520863, gi:20520869.
FEATURES             Location/Qualifiers
     source            1..295150
                     /organism="Streptomyces coelicolor A3(2)"
                     /mol_type="genomic DNA"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     512..515
RBS                   /note="possible RBS upstream of SC7B7.02"
                     524..1933
gene                  /gene="SC06005"
                     /note="synonym: SC7B7.02"
                     524..1933
CDS                   /gene="SC06005"
                     /note="SC7B7.02, possible lipoprotein, len: 469; has
                     similarity to signal sequence of Mycobacterium leprae
                     hypothetical protein TR:E334742 (EMBL:298604) MLCB2052.27
                     (445 aa), fasta scores; Opt: 75 z-score: 199.2 E():
                     0.00062, 23.3% identity in 335 aa overlap. Contains
                     N-terminal signal sequence and appropriately positioned
                     PS00013 Prokaryotic membrane lipoprotein lipid attachment
                     site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="CAA15782.1"
                     /db_xref="GI:2661678"
                     /db_xref="GOA:O50499"
                     /db_xref="SPTREMBL:O50499"
                     /translation="MTIRAGSLDRITLLGAIATAMGSAFVACSPSSDKSDSGP
                     KGEKSNPFAGANSTVEAAIFDGYGTVDYDANQVLSQVKGKLVQVQPVVDIAP
                     QLOPRFVGNPDLIDNSGDOIIGLIGLDLELDFEASTVEGKKIADIIVPGVK
                     DPGFKDKVNAVNVYGVYKTLFEENGWTPPKTWDEALDLGQAAKGGKILFV
                     HGKAARYRLLIDSLKGGDEVRLEALEKGCWSPAVQGVKMWETVYKQKMF
                     VPGSGTQFQAQAIWINDQKALLPSCGHIEEMKATKADQMGIGIPSWITLDPKA
                     LPYALRAKQEPFIVPQGNPAGKEVRLMLSEKAAANFKTKLPIATVIGTVPA
                     DGYSTALVSTQKLEAAGTIFNFMFVETVGLNTDQLVPWNSFLAGLDLKGKGLTSAL
                     QKISDKVREDDSDVKVYS"
                     581..613
misc_feature          /gene="SC06005"
                     /note="PS00013 Prokaryotic membrane lipoprotein lipid
                     attachment site"
                     1941..3002
gene                  /gene="SC06006"
                     /note="synonym: SC7B7.03"
                     1941..3002
CDS                   /gene="SC06006"
                     /note="SC7B7.03, probable integral membrane transport
                     protein, len: 353 aa, similar to many members of the MALFG
                     subfamily of the binding-protein-dependent transport
                     system membrane proteins eg. MSMP_STRMU_Q00750 multiple
                     sugar-binding transport system permease protein (290 aa),
                     fasta scores; Opt: 393 z-score: 636.8 E(): 2.6e-28, 32.5%
                     identity in 280 aa overlap. Contains PS00402
                     Binding-protein-dependent transport systems inner membrane
                     comp signature"
                     /codon_start=1
                     /transl_table=11
                     /product="probable transmembrane transport protein"
                     /protein_id="CAA15783.1"
                     /db_xref="GI:2661678"
                     /db_xref="GOA:O50500"
                     /db_xref="SPTREMBL:O50500"
                     /translation="MKDIPTAETASRPPEAAGGRPRRKLIFDRVTFFLAFLGVP
                     LAIFVFLIPGQAIFWGTDWRFSPDINFGVDFNFTKFDQDIFPLKUNVALLLA
                     AFVPLVTLALGVAITLGGSPKGPVRIGASFYRIISFFYPVPAIIVNALLQA
                     MYDPNAGLNGVLTGLGDQFOTFAWLGEKAAMPANFVIMVGLVGFYVITFAAIK
                     GVPGLYEAAKIDGAGRTTISITLPAIRDSVOTAYIYLGIALDAPFYVQAVPVG
                     GPDNSTLTISQRLFNVAFAKQFGVATANGVVLAAVTLVFAALFVLNRLTGGCGEGES
                     KKAQGSRRBBAAKGAR"
                     2589..2675
misc_feature

```

```

/gene="SC06006"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane component signature"
2988..2992
/gene="SC06006"
/note="possible RBS upstream of SC7B7.04"
2999..3907
/gene="SC06007"
/note="synonym: SC7B7.04"
2999..3907
/gene="SC06007"
/note="SC06007"
2999..3907
/note="SC06007"
/note="SC7B7.04, probable integral membrane transport
protein, len: 301 aa; similar to many members of the MALFG
subfamily of the binding-protein-dependent transport
systems membrane proteins eg. TR:Q06111 (EMBL:X66092)
putative transport system permease protein from
Clostridium perfringens (275 aa), fasta scores; Opt: 590
z-score: 796.5 E(): 0, 33.6% identity in 265 aa overlap.
Contains PS00402 Binding-protein-dependent transport
systems inner membrane comp signature"
/codon_start=1
/transl_table=11
/product="probable transmembrane transport protein"
/protein_id="CAA15784.1"
/db_xref="GI:2661679"
/db_xref="GOA:O50501"
/db_xref="SPTREMBL:O50501"
/translation="MSVIAADKPKASDRKPVSDRKLAPAAASDRRFAAISHALLILW
SVITVPMVLMSCVSKFGTBIILSPFSLPDHWFENYANAWTDANIKGYFLNSVIIV
VSALLVLMLGAMCAVLAIRFEFQRRLLIYVMLAGLTFPVPLAIVPFOIQNFGLL
NTRFGLIITYAPALPTFMFLYSFERSLPHDVVEAALIDGAGDWRAFPQVMLPWARP
GNAVAIFNYGLMNOFLPVALNIDQKWLVTQGMAYASSQYVDIDYGALFAAIVV
TVPVLLVYCVFQRRIGSVSQGTR"
3554..3640
/gene="SC06007"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp signature"
3935..3984
/note="hairpin loop with 23bp stem"
4128..4131
/note="possible RBS upstream of SC7B7.05"
4142..5353
/gene="SC06008"
/note="synonym: SC7B7.05"
4142..5353
/gene="SC06008"
/note="SC7B7.05, probable transcriptional repressor
protein, len 403 aa; similar to many eg. XYLX_AARH_Q44406
xylose repressor from Anaerocellum thermophilum (399 aa),
fasta scores; Opt: 548 z-score: 767.3 E(): 0, 33.1%
identity in 393 aa overlap. Contains PS01125 ROK family
signature and helix-turn-helix motif at aa positions 29-50
(score 2016, +6.05 SD)"
/codon_start=1
/transl_table=11
/product="probable transcriptional repressor protein"
/protein_id="CAA15785.1"
/db_xref="GI:2661680"
/db_xref="GOA:O50502"
/db_xref="SPTREMBL:O50502"
/translation="METFGQSLSLHRLNLRVVRVAVRAGSLTQAEIARTTGLSAAATV
SNIRLEKGGTVETVPTTSAGGRKARSVLSGDAGIVIGVDGFGHRLRVAVGNLAHOV
LABESPLDQVTSAGQDFRQALVRLIEATGVDRAKIAVGGLGVPFIDVSGTILG
STAALFGWGIIRPABELRGRLGVPVHNDNANGALGELVWGGGGRVRLATYIKVASG
VGAGLVNGKIYRGQTAGETAGHITLDEAGFVRCGNRGCLTETFAAFYVLPILQPS
HGTDLMGVVLRVARDPDGCRVADVGRHIGSGVNLCLNLSRVRVVLGDLAEAG
EALVGPIRSVGRVYAIIPAARQLSVLPGLAGRAEVLGALALSEMGGDSTLLDGSAT
GALPAATPAFT"
4784..4867
/gene="SC06008"
/note="PS01125 ROK family signature"
5462..5466
/note="possible RBS upstream of SC7B7.06"

```

```
gene      5475..6590
/ gene="SCO6009"
/ note="synonym: SC7B7.06"
CDS       5475..6590
/ gene="SCO6009"
/ note="SC7B7.06, probable solute-binding protein of
transmembrane transport system, len: 403 aa; similar to
many eg. XYL ECOLI P37387 d-xylose-binding periplasmic
protein precursor (330 aa), fasta scores; Opt: 619
z-score: 573.4 E(): 8.9e-25, 35.7% identity in 325 aa
overlap. Contains N-terminal signal sequence and
appropriately positioned PS00013 Prokaryotic membrane
lipoprotein lipid attachment site"
/codon_start=1
/transl_table=11
/product="solute-binding protein"
/protein_id="CAA15786.1"
/db_xref="GI:2861661"
/translation="NNTMRRAAVALTAMAVSLACGSAKESDGGSDSDSAKKG
DDIKVGLLLPENQYARVEKFKPMIEKKVKELTNKGEVVYANAKQDASLQNOQVDITM
Query Match      12.8%; Score 64; DB 1; Length 295150;
Best Local Similarity 51.5%; Pred. No. 3.8; Indels 6; Gaps 1;
Matches 176; Conservative 0; Mismatches 160;
QY 141 CACGTGTACGCGGGAGCTACAGGATCCACTGTCCGCCAGCCCGCGGAACTCCTGCTC 200
Db 187772 CAGGTGCAGCGGTTCCGCGCAGCGGTCAGGATCCGTGCTCGAGGAACGCTCCAG 187831
QY 201 CTTGCCATGATCTCGTCGGCGTGGTTCAGCGGAGAGCAGCGGTAGTCCACCGCTC 260
Db 187832 CGTCGGGATCAGTCGCGCGGTCGGCGGTCGTAGTCCCGCAGCGGTCAGCAGCG 187891
QY 261 GGGCGTGAACGGCTCCCGGGTGCACACCGGATGTGCGTCCCGGGGTGAGCGGCCCTG 320
Db 187892 GCGGTTGAAGCGCGCGGACGTCGTCGGGACGAA-----GGGAGCAGCAGCACGTG 187945
QY 321 CTTGCCGCGTGTGTGTCACACCGAGAGACAGTCCGAGACCGATCCCGCAGAGTT 380
Db 187946 CGAGGCCAGTCTCTGTGCGCGCCCGCGCAGACCCCGCGGGCGCGCGCGGACCCG 188005
QY 381 CGTCACGGTGGCGCTCTTCGCGTCCGCGCGTACGCCACCCCGCTTGCCCTCGCCCTT 440
Db 188006 GCGCGGTGCGCGGCTCTTCAGCGCCCGCGCAGCCCTTCGCGAGTGCACGCGCGC 188065
QY 441 GAGCGAGTTCAGAGGCGCAGAGGTGCGTGCAGATGCCCTC 482
Db 188066 GCTGACCGGAGCGTCACGCGCCGCTCGCCGCTCCAGGCCCGC 188107
```

Search completed: June 27, 2004, 19:39:06  
Job time : 2371.22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:13 ; Search time 334.804 Seconds  
(without alignments)  
6841.451 Million cell updates/sec

Title: US-09-758-759-1 COPY 1 500

Perfect score: 500  
Sequence: 1 ggtaccgcgacccgtgtccgg .....tcgacgtcggcggaacct 500

Scoring table: IDENTITY NUC

Gapex 10.0, Gapext 1.0

Searched: 3017426 segs. 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

**POST-PROCESSING: MINIMUM MATCH 0%      Maximum Match 100%**

Maximum Match 100%  
Listing first 50 summaries

```

Database :
Published Applications_NA:**
1: /csm2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:**
2: /csm2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:**
3: /csm2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:**
4: /csm2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:**
5: /csm2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:**
6: /csm2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:**
7: /csm2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:**
8: /csm2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:**
9: /csm2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:**
10: /csm2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:**
11: /csm2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:**
12: /csm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:**
13: /csm2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:**
14: /csm2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:**
15: /csm2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:**
16: /csm2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:**
17: /csm2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:**
18: /csm2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:**
19: /csm2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:**

```

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
C 1	500	100.0	37116	13	US-10-107-431-279	Sequence 279, App
C 2	500	100.0	109519	12	US-09-758-759-1	Sequence 1, Appli
C 3	369	73.8	1251	13	US-10-107-431-232	Sequence 232, App
C 4	366	73.2	1248	12	US-09-758-759-2	Sequence 2, Appli
C 5	298.4	59.7	10035	13	US-10-107-431-282	Sequence 282, App
C 6	298.4	59.7	11115	10	US-09-769-734-49	Sequence 49, Appl
C 7	256.6	51.3	1251	13	US-10-107-431-230	Sequence 230, App
C 8	131	26.2	12152	12	US-09-758-759-182	Sequence 182, App
C 9	124	24.8	603	13	US-10-107-431-220	Sequence 220, App
C 10	118.2	23.6	15240	13	US-10-107-431-278	Sequence 278, App
C 11	112.8	22.6	1236	12	US-09-758-759-50	Sequence 50, Appl
C 12	112.8	22.6	1281	13	US-10-107-431-122	Sequence 122, App
C 13	97	19.4	63158	15	US-10-232-198-1	Sequence 1, Appli
C 14	92.4	18.5	12390	13	US-10-107-431-120	Sequence 120, App

## ALIGNMENTS

## RESULT 1

US-10-107-431-279/c

US 10,107,431 ; Sequence 279, Application US/10107431

Publication No. US20030224364A1

: GENERAL INFORMATION:

APPLICANT: Garnet

```

; APPLICANT: Farnet,
: APPLICANT: Staffa

```

APPLICANT: STAFF, ZAZONO

APPLICANT: ZAZOPOL  
TITLE OF INVENTION:

; TITLE OF INVENTION:

; TITLE OF INVENTION:

FILE REFERENCE: 300

**; CURRENT APPLICATION**

; CURRENT FILING DATE

; NUMBER OF SEQ ID NO

; SOFTWARE: PatentIn

; SEQ ID NO 279

LENGTH: 37116

LENGTH: 3,110  
TYPE: DNA

Query Match 100.0%: Score 500: DB 13: Length 37116:

Query Match	100.0%;	score 300;	DB 13;	length 37118;
Best Local Similarity	100.0%;	Pred. No. 2.4e-113;		

Best Local Similarity 100.0%; Fied. NO. 2.4E-113;  
Matches 500. Conservative 0. Mismatches 0. Indels 0. Gaps 0.

0-1

QY 1 GGTACCCGACCGTGTCTCCGGAAACAACGAGTCCGAGATACGGCGAGAGGAACACCCCGGT 60

Db 30486 GGTACCCGACCGTGTCCTCCGGAAACAACGAGTCGAGATACGGCGAGAGGAACACCCCCCGGT 3042

[illegible]

QY 121 ACACCGGGGTGAGTCTGCTACCTGTACCGGGGGAGCTACAGGATCCACTGTCCGCCA 180  
Db 30366 ACACCGGGGTGAGTCTGCTACCTGTACCGGGGGAGCTACAGGATCCACTGTCCGCCA 30307  
QY 181 GCCCGGGGAACTCTGCTCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 240  
Db 30306 GCCCGGGGAACTCTGCTCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 30247  
QY 241 AGCCGCTAGTCCACCGGCTGGGGGTGAAACCGCTCCGGGGTGCACACCGGGATGTGGTG 300  
Db 30246 AGCCGCTAGTCCACCGGCTGGGGGTGAAACCGCTCCGGGGTGCACACCGGGATGTGGTG 30187  
QY 301 CGGGGGGTAGCGGGCCCTGCTTGGCGGGCTGCTGTCGACACACCGAGGAGACAGGTCC 360  
Db 30186 CGGGGGGTAGCGGGCCCTGCTTGGCGGGCTGCTGTCGACACACCGAGGAGACAGGTCC 30127  
QY 361 GGACCGATGCCCGCAAGATTGCTACGGTGGCGCTCTTTCGCCGTGCGCCGCTAGCCACC 420  
Db 30126 GGACCGATGCCCGCAAGATTGCTACGGTGGCGCTCTTTCGCCGTGCGCCGCTAGCCACC 30067  
QY 421 ACCCGCTTGCCTCGGCGCTTGCAGGATTCAGCGGGGAGCAGGTGCGTGGATGCC 480  
Db 30066 ACCCGCTTGCCTCGGCGCTTGCAGGATTCAGCGGGGAGCAGGTGCGTGGATGCC 30007  
QY 481 TCGACGTGCGGCGGAACCT 500  
Db 30006 TCGACGTGCGGCGGAACCT 29987  
RESULT 2  
US-09-758-759-1  
; Sequence 1, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Evernimicin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 109519  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
US-09-758-759-1  
Query Match 100.0%; Score 500; DB 12; Length 109519;  
Best Local Similarity 100.0%; Pred. No. 2e-113;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGGTGTCGCCGAAACAGAGTTCGAGATACCGGCGAGAGAACACCCCGGT 60  
Db 1 GGTACCCGACCGGTGTCGCCGAAACAGAGTTCGAGATACCGGCGAGAGAACACCCCGGT 60  
QY 61 AGTCCGGGTAGACGGGTGGCGGCGAAGCGTACCGGCTTCGACCGGTACCGGCGGGCGGG 120  
Db 61 AGTCCGGGTAGACGGGTGGCGGCGAAGCGTACCGGCTTCGACCGGTACCGGCGGGCGGG 120  
QY 121 ACACCGGGGTGAGTCTGCTACCTGTACCGGGGGAGCTACAGGATCCACTGTCCGCCA 180  
Db 121 ACACCGGGGTGAGTCTGCTACCTGTACCGGGGGAGCTACAGGATCCACTGTCCGCCA 180  
QY 181 GCCCGGGGAACTCTGCTCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 240  
Db 181 GCCCGGGGAACTCTGCTCTTCCGCAATGATCTGCTCGGGGTGTTTCCAGGCGAAGAGC 240  
QY 241 AGCCGCTAGTCCACCGGCTGGGGGTGAAACCGCTCCGGGGTGCACACCGGGATGTGGTG 300

Db 241 AGCCGCTAGTCCACCGGCTGGGGGTGAAACCGCTCCGGGGTGCACACCGGGATGTGGTG 300  
QY 301 CGGGGGGTGAGCGGGCCCTGCTTGGCCCGGCTGCTGTCGCACACCGAGAGACAGGTCC 360  
Db 301 CGGGGGGTGAGCGGGCCCTGCTTGGCCCGGCTGCTGTCGCACACCGAGAGACAGGTCC 360  
QY 361 GGACCGATGCCCGCAAGATTGCTACGGTGGCGCTCTTTCGCCGTGCGCCGCTAGCCACC 420  
Db 361 GGACCGATGCCCGCAAGATTGCTACGGTGGCGCTCTTTCGCCGTGCGCCGCTAGCCACC 420  
QY 421 ACCCGCTTGCCTCGGCGCTTGCAGGATTCAGCAGGGGAGCAGGTGCGTGGATGCC 480  
Db 421 ACCCGCTTGCCTCGGCGCTTGCAGGATTCAGCAGGGGAGCAGGTGCGTGGATGCC 480  
QY 481 TCGACGTGCGGCGGAACCT 500  
Db 481 TCGACGTGCGGCGGAACCT 500  
RESULT 3  
US-10-107-431-232/c  
; Sequence 232, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTH  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 232  
; LENGTH: 1251  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea africana  
US-10-107-431-232  
Query Match 73.8%; Score 369; DB 13; Length 1251;  
Best Local Similarity 100.0%; Pred. No. 4.1e-81;  
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 132 TCAGCTCGTCACTGTCAGCGGGGAGCTACAGATCCACTGTCGCCAGCCCGCGGAA 191  
Db 1251 TCAGCTCGTCACTGTCAGCGGGGAGCTACAGATCCACTGTCGCCAGCCCGCGGAA 1192  
QY 192 CTCCTGCTCTTCGCCATGATCTCGTGGCGTGGTTCAGGCGAAGAGCAGCGGTAGTC 251  
Db 1191 CTCCTGCTCTTCGCCATGATCTCGTGGCGTGGTTCAGGCGAAGAGCAGCGGTAGTC 1132  
QY 252 CACCGCTGGGGGTGAAACCGCTCCGGGGTGCACCGGATGTCGTGCGGGGGGTGAG 311  
Db 1131 CACCGCTGGGGGTGAAACCGCTCCGGGGTGCACCGGATGTCGTGCGGGGGGTGAG 1072  
QY 312 CCGGCCCTGCTTGGCGCGGTGCTGTCGCACACCCAGGAGACAGGTCCGGAACGATGCC 371  
Db 1071 CCGGCCCTGCTTGGCGCGGTGCTGTCGCACACCCAGGAGACAGGTCCGGAACGATGCC 1012  
QY 372 GCAGAGTTGCTACCGTGGCGCTCTTCGCCGTGCGGCGCTAGCCACACCGCTTGGC 431  
Db 1011 GCAGAGTTGCTACCGTGGCGCTCTTCGCCGTGCGGCGCTAGCCACACCGCTTGGC 952  
QY 432 CTCGGCTTGCAGCGATTTCAGCAGGGGAGCAGGTGCGTGGATGCCCTTCGACGTGGC 491  
Db 951 CTCGGCTTGCAGCGATTTCAGCAGGGGAGCAGGTGCGTGGATGCCCTTCGACGTGGC 892  
QY 492 GCGGAACCT 500  
Db 891 GCGGAACCT 883

```
RESULT 4
US-09-758-759-2/c
; Sequence 2, Application US/09758759
; Publication No. US20040101832A1
; GENERAL INFORMATION:
; APPLICANT: Hosted, Thomas J.
; APPLICANT: Wang, Tim X.
; APPLICANT: Horan, Ann C.
; TITLE OF INVENTION: Evernimicin Biosynthetic Genes
; FILE REFERENCE: ID0983K US
; CURRENT APPLICATION NUMBER: US/09/758,759
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: US 60/175,751
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1248)
; OTHER INFORMATION: evda
US-09-758-759-2

Query Match 73.2%; Score 366; DB 12; Length 1248;
Best Local Similarity 100.0%; Pred. No. 2.2e-80;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 GCTGTCACGTGTACGCGGGGACGTACAGGATCCACTGTGCGCCAGCCGCGGAATC 194
DB 1248 GCTGTCACGTGTACGCGGGGACGTACAGGATCCACTGTGCGCCAGCCGCGGAATC 1189

QY 195 CTGCTCTTCGCATGATCTCTGCGGTGTCGACCGGGATGTGCGTCCCGGGGTGAGCGG 314
DB 1188 CTGCTCTTCGCATGATCTCTGCGGTGTCGACCGGGATGTGCGTCCCGGGGTGAGCGG 1069

QY 315 GCCTGCTTGGCGGGGTCTGTGCGACACCGGAGACAGAGTCCGACCGATGCCGA 374
DB 1068 GCCTGCTTGGCGGGGTCTGTGCGACACCGGAGACAGAGTCCGACCGATGCCGA 1009

QY 375 GAAGTTGCTACAGGTGGCGCTCTTCCGCGTCCGCGGTACGCCACCGCTTGCCTC 434
DB 1008 GAAGTTGCTACAGGTGGCGCTCTTCCGCGTCCGCGGTACGCCACCGCTTGCCTC 949

QY 435 GGCTTTGAGCGAGTTACGAGGGGACAGGTGCGGTGCGGATGCCCTCGACGTCGCGGC 494
DB 948 GGCTTTGAGCGAGTTACGAGGGGACAGGTGCGGTGCGGATGCCCTCGACGTCGCGGC 889

QY 495 GAACCT 500
DB 888 GAACCT 883

RESULT 5
US-10-107-431-282/c
; Sequence 282, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; FILE REFERENCE: BIOSYNTHETIC LOCI
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
```

```
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 282
; LENGTH: 10035
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea aurantiaca
US-10-107-431-282

Query Match 59.7%; Score 298.4; DB 13; Length 10035;
Best Local Similarity 76.1%; Pred. No. 5e-64;
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;

QY 1 GGTACCCGACCGTGTCCCGGAACACGAGTCAGATACGCGGAGAGAACACCCCGGT 60
DB 3742 GATATCCGAGCGTCTCGGTGAAGACCGAGTCCAGTACGGGACACAGAGACCCCGC 3683

QY 61 AGTCCGGGTACACGCTGGGCGCGAAGCGTACGCGCTTCCAGCGTCAGCGGGCGGCG 120
DB 3682 GGTCCGGGAAGACCCCGGGGTGAAGACGTACGCGCCCTCGACGCGGAGTTCGGTTGCTA 3623

QY 121 ACACC---GGCGGGTCACTCGTCACTGTCAGTGTACGGGGGAGCTACAGGATCCACTGCC 176
DB 3622 CTTGACAGGCCCTCACTCAATCCCGACGTGAACGTGCGGAACGTACAGGATCCAGGCCCC 3563

QY 177 GCCAGCCCGCGGAACTCTCTCTTCGCCATGATCTCGTGGCGTGGTTCAGAGCGAA 236
DB 3562 GCCGCGCTGCGGAAAGCGCTCTCTTCGCCATGATCTCGTGGCGTGGTTCAGAGCGAA 3503

QY 237 GAGCAGCGGTACTCCACCGCGTCCGGGCTGAACGCGTCCGGGTGCGCACCGGGATGTG 296
DB 3502 GAGCAGCGGTACTCCACCGCGTCCGGGCTGAACGCGTCCGGGTGCGCACCGGGATGTG 3443

QY 297 CGTCCCGGGGTGAGCGCGCCCTCTTGGCGCGGCTCGTTCGCGACACACCCAGAGACCG 356
DB 3442 CGTCCCGGGGTGAGCGCGCCCTCTTGGCGCGGCTCGTTCGCGACACACCCAGAGACCG 3383

QY 357 GTCGCGACCGATGCCGCGAAGTTCGTCACGCTGGCGCTCTTCCGCGTCCGCGCGTACGC 416
DB 3382 GTCGCGACCGATGCCGCGAAGTTCGTCACGCTGGCGCTCTTCCGCGTCCGCGCGTACGC 3323

QY 417 CACACCGCGTTCCTCGCGCTTTCGCGAGTTCAGCAGGCGGAGCAGGTGCGTCCGAT 476
DB 3322 CACACCGCGTTCCTCGCGCTTTCGCGAGTTCAGCAGGCGGAGCAGGTGCGTCCGAT 476

QY 477 GCGTTCGACGTGCGCGGCGAAC 498
DB 3262 CCGCGCGACCTGGCGAGCGAAC 3241

RESULT 6
US-09-769-734-49/c
; Sequence 49, Application US/09769734
; Publication No. US2003014366A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; TITLE OF INVENTION: Genetic Locus for Evernimicin Biosynthesis
; FILE REFERENCE: PA 005-US
; CURRENT APPLICATION NUMBER: US/09/769,734
; CURRENT FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 49
; LENGTH: 11115
; TYPE: DNA
; ORGANISM: M. carbonacea
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (8)..(1207)
; OTHER INFORMATION: ORF 41 (positive strandedness)
; OTHER INFORMATION: incomplete: C-terminus only
; NAME/KEY: misc.feature
; LOCATION: (1213)..(2331)
; OTHER INFORMATION: ORF 42 (positive strandedness)
```

NAME/KEY: misc feature  
LOCATION: (2364)..(3611)  
OTHER INFORMATION: ORF 43 (positive strandedness)  
NAME/KEY: misc feature  
LOCATION: (3623)..(4243)  
OTHER INFORMATION: ORF 44 (positive strandedness)  
NAME/KEY: misc feature  
LOCATION: (4149)..(5177)  
OTHER INFORMATION: ORF 45 (positive strandedness)  
NAME/KEY: misc feature  
LOCATION: (5177)..(6094)  
OTHER INFORMATION: ORF 46 (negative strandedness)  
NAME/KEY: misc feature  
LOCATION: (6271)..(7824)  
OTHER INFORMATION: ORF 47 (negative strandedness)  
NAME/KEY: misc feature  
LOCATION: (7903)..(8760)  
OTHER INFORMATION: ORF 48 (negative strandedness)  
NAME/KEY: misc feature  
LOCATION: (8781)..(9800)  
OTHER INFORMATION: ORF 49 (negative strandedness)  
US-09-759-734-49

Query Match 59.7%; Score 298.4; DB 10; Length 11115;  
Best Local Similarity 76.1%; Pred. No. 4.9e-64;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
QY 1 GGTACCCGACCGTGTCCCGGACACGAGTCAGATACGGGAGAGAACACCCCGGT 60  
DB 3749 GATATCCGAGCGTCTCGGTGAGACCGAGTCAGATACGGGAGACACCCCGCG 3690  
QY 61 AGTCGCGGTAGACGCGTGGCGGCGAAGCGTACCGGCCCTTCGACGCGTCAGCGGGCGGCGG 120  
DB 3689 GGTGGGAAAGACCCGCGGGGTGAAGACGTACGGGCCCTTCGACGCGGAGTTCGGTTGTA 3630  
QY 121 ACACC-----GGCGGTACGTCTGTCAGTGTACGGGGGAGTACAGGATCCACTGCC 176  
DB 3629 CCGTACGCGGCCCACTCAATCCGACGTGAACGTCGCGGAACGTACAGGATCCAGGCCCC 3570  
QY 177 GCACGCGCGGGGAACTCTGCTCTCCGATGATCTGTCGCGGTGTTCCAGGCGAA 236  
DB 3569 GCCGCGCTGCCGGAACGCTGCTCTTCGCGATGATCTGCGGTGTTCCAGCGAA 3510  
QY 237 GAGCAGCGGTAGTCCACGCGGTGCGGGGTGAACGCGTCCGGGTGCGCACCGGATGTG 296  
DB 3509 GAGCAGCGGTAGTCCGCGGGTGGTGGGAATCGTCCCGCGATGAACCGGATGTG 3450  
QY 297 CGTCCGCGGGTGAGCGCGCCCTGCTTGGCGCGGTGTCGTCGCGGTGTTCCAGGCGAA 356  
DB 3449 CGTCCGCGGGTACGCGCGCCCTGTTTGGCGGGTCTGTCGTACACCCCGCACACAG 3390  
QY 357 GTCGCGACCGATGCCGCGAAGTTCGTACGCGTGGCGTCTTCCGCGTGGCGCGTACGC 416  
DB 3389 GTCCGCGCGGATGCCGCGAAGTTCGCCACGCTCGCGTCTTGGCGGTGCGCCCGTAGGC 3330  
QY 417 CACACCGCGTTCCTCGGCTTCAGCGAGTTCAGCAGGCGAGCAGGTGCGGTGCGAT 476  
DB 3329 CACGACCGGTTTGCCTTCGCGGTCAACCGTTCGAGCAGCCCGCAGGTGTCGCGAT 3270  
QY 477 GCGCTCAGCTCGCGGGGAAC 498  
DB 3269 CCGGCGGACCTGGGAGCGAAC 3248

RESULT 7  
US-10-107-431-230/c  
; Sequence 230, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO

TITLE OF INVENTION: BIOSYNTHETIC LOCI  
FILE REFERENCE: 3001-705  
CURRENT APPLICATION NUMBER: US/10/107,431  
CURRENT FILING DATE: 2002-03-28  
NUMBER OF SEQ ID NOS: 282  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 230  
LENGTH: 1251  
TYPE: DNA  
ORGANISM: Micromonospora carbonacea aurantiaca  
US-10-107-431-230  
Query Match 51.3%; Score 256.6; DB 13; Length 1251;  
Best Local Similarity 81.2%; Pred. No. 1.1e-53;  
Matches 298; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 132 TCAGCTCGTCAGCTGTACGCGGGGACGTACAGGATCCACTGTCCGCGACCCCGCGGAA 191  
DB 1251 TCAATCCCGCACGTGAACGTGCGGAACGTACAGGATCCAGGCCCGCGGCTGCCGAA 1192  
QY 192 CTCCTGCTCTTCGCCATGATCTCGTCGCGGTGTTCCAGGCCAAGAGACGCGGTAGTC 251  
DB 1191 CGCTGCTCTTCGCCATGATCTCGTCGCGGTGTTCCAGGCCAAGAGACGCGGTAGTC 1132  
QY 252 CACCGCGTCGCGGTGAACGCGTCGCGGGTGGCGACCGGGATGTCGTCGCGGGGTAG 311  
DB 1131 CGCGCGGTGCGTCGCGGAACCTCGTCGCGCGCATGAACCGGGATGTCGTCGCGGGTCA 1072  
QY 312 CGCGCGCTCTTCGCGCGGTGTCGTCGACACCCAGGAGACGAGTCCGAGCCGATGCC 371  
DB 1071 GCGCGCTCTTCGCGCGGTGTCGTCGACACCCGAGACACGAGTCCGCGCGGTAGCC 1012  
QY 372 GCAGAGATTCGTACGCGTGGCGCTCTTCGCGTGGCGGTACGCGACACCCGCTTGCC 431  
DB 1011 GCAGAGATTCGCGACGCTCGCGCTCTTCGCGGTGCGCGGTAGGCGCACCCGCTTGCC 952  
QY 432 CTCGCGCTTACGCGATTCAGCAGGCGGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGG 491  
DB 951 CTCGCGCTTACGCGATTCAGCAGGCGGAGCGGTGCGGTGCGGTGCGGTGCGGTGCGG 892  
QY 492 GCGGAAC 498  
DB 891 AGCGAAC 885  
RESULT 8  
US-09-758-759-182  
; Sequence 182, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Evernimicin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 182  
; LENGTH: 12152  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
US-09-758-759-182  
Query Match 26.2%; Score 131; DB 12; Length 12152;  
Best Local Similarity 100.0%; Pred. No. 3.6e-23;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGTGTCCCGGAAACAGAGTCGAGATACGCGGAGAGAACACCCCGGT 60



Db 12022 GGTACCCGACCGTGTCCCGGAAACAACAGAGTCGAGATACGGCGAGAGGAACAACCCCGCGGT 12081  
QY 61 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 120  
Db 12082 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 12141  
QY 121 ACACCGCGGG 131  
Db 12142 ACACCGCGGG 12152

RESULT 9

US-10-107-431-220/c  
; Sequence 220, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-705  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 220  
; LENGTH: 603  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea africana  
US-10-107-431-220

Query Match 24.8%; Score 124; DB 13; Length 603;  
Best Local Similarity 100.0%; Pred. No. 3e-21;  
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACCCGACCGTGTCCCGGAAACAACAGAGTCGAGATACGGCGAGAGGAACAACCCCGCGGT 60  
Db 124 GGTACCCGACCGTGTCCCGGAAACAACAGAGTCGAGATACGGCGAGAGGAACAACCCCGCGGT 65  
QY 61 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 120  
Db 64 AGTCCGGGTAGACGGTGGCGCGGCGAAGCGGTACCGCGCTTCGACGGTCAAGCGGCGGCGG 5  
QY 121 ACAC 124  
Db 4 ACAC 1

RESULT 10

US-10-107-431-278/c  
; Sequence 278, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-705  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 278  
; LENGTH: 15240  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea africana  
US-10-107-431-278

Query Match 23.6%; Score 118.2; DB 13; Length 15240;  
Best Local Similarity 53.3%; Pred. No. 4.6e-20;  
Matches 249; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 33 AGATACCGCGAGAGGAACAACCCCGGTAGTTCGGGTAGACGGTGGCGGCGAAGAGCGGTAC 92  
12684 AGAGATGATCCAGCGGAGCAGCAGCATGGTCCGGGACCGGGCTTCGGGGTCAATGA 12625  
QY 93 GCGCTTCGACGGTCAGCGGCGGCGGACACCGCGGGTCAGTCTCGTACGTTAGCGG 152  
Db 12624 AAGTTGATCAGTCTCGGTGGGGCGGTGTATCCCGGTCCGCGGACCGCTCTCGACCGG 12565  
QY 153 GGGACGTPACAGGATCCACTGTCCGCGAGCCCGCGGAACTCTGTCTCTTCGCGCATGAT 212  
Db 12564 GGGCAGCGGAAAGATCAACCGGTGCGCGTGGCGAGCATCTCCGCTCTCCGCGGACGAT 12505  
QY 213 CTCGTCCGCGTGGTTCAGAGCGAAGAGCAGCGGTAGTCCACCGCGTCCGCGGTGAACGC 272  
Db 12504 CTCGTCCGGAAGTGCAGGCGAGCAGCATAGTAGTATCCGGGCGGCGCGCGCATC 12445  
QY 273 GTCGGGGTGCACACCGGGATGTGCTGCCGGGGGTAGCGCGCCCTGTGTCGCGGGCGGT 332  
Db 12444 CTGCTCGCTGATGATTTTCGATGTCGTCGAGGGTGCAGCGCCACCTTGTCCGGATT 12385  
QY 333 GGTGTCGACACACCGAGGACCGGTCCGAGCGATGCGGAGTTCGTCACGCTGGC 392  
Db 12384 GCGCTCCGCGGTACCGGATGAGTCCGGTCTATGCCGCGAGAACTGCAGAGGGTGT 12325  
QY 393 GCTCTTCGCGTCCGCGCGGTACGCGCACCCCGCTTGCCTTCGCGCTTGACGAGTTCAAG 452  
Db 12324 GCGCTTCGTCGAGCGCGGTAGAGTGCACGTCGCGGCGCTTGGCGCGCGAGCTCGCGCAG 12265  
QY 453 CAGGCGGAGAGTCCGTGCGGATGCCCTTCGACGTCGCGGCGGCGGAC 499  
Db 12264 CAGGCGGCTCACCTCGTCACCGTGTGCGCGCACCTGTCTCGCGGAGC 12218

RESULT 11  
US-09-758-759-50/c  
; Sequence 50, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Everminomycin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 50  
; LENGTH: 1236  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1236)  
; OTHER INFORMATION: evm  
US-09-758-759-50

Query Match 22.6%; Score 112.8; DB 12; Length 1236;  
Best Local Similarity 56.5%; Pred. No. 1.5e-18;  
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 128 CGGTTCAGTCTCGTACCGGGGAGCGTACAGGATCCACTGTCCGCGGCGGCGGCGG 187  
Db 1234 CGGTTCGCGCGCGGTCTCGACCGGGGAGCGGAAAGATCAACCGGGTCCGCTGGCCA 1175  
QY 188 GGAATCTCTCTCTCTTCGCGCATGATCTCGTCCGCGTTCAGGCGGAAAGAGAGCGGT 247  
Db 1174 GCATCTCCGCTCCCGGGGAGCATCTCGTCCCGGAAAGTCCGAGGAGGACGAGTACT 1115  
QY 248 AGTCCACCGCGTCCGGCGGTGAACGCGTCCGGGGTGGCACCGGGATGTCTCGTCCGCGGCGG 307



```
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea aurantiaca
US-10-107-431-120

Query Match.      18.5%; Score 92.4; DB 13; Length 1290;
Best Local Similarity 53.6%; Pred. No. 1.4e-13;
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 142 ACCTGTACGCGGGGAGCTACAGGATCCACTGTCCGCCAGCCCGGGGAACTCTCTCTCC 201
DB 1250 ACCCGCAGGCTGGGCGAGGAGATCAGCTTGTTCCGSCGCCACCGTGGCGCTCG 1191
QY 202 TTGCGCATGATCTCGTCGGCGTGTTCAGGCGAAGAGCAGCGCGTAGTCCACCGGTG 261
DB 1190 CGCGCCACGATCTCGTCGTGGAAGTGCACCGCAGCACCAGGTAGTGGTCCGGCGCG 1131
QY 262 GGGGTGAACCGCTCCGGGGTGGCGCACCGGATGTGGCTGCCGGGGGTGAGCCGGCCCTGC 321
DB 1130 GCCCGCAGTCCGCTCGCTGATCATCTCGATGTCCGTACCGAGGTCCGGCGCGACC 1071
QY 322 TTGCGCGCGTGTGTGCGACACCCAGGACACCGAGTCCCGACCGATGCGCGAAGATTC 381
DB 1070 TTGTCCGGTGTGCGTCCGGCGGTACGGATCAGCGTGGTGGTCCGCGAGTACTGC 1011
QY 382 GTACAGTGTGGCTCTTCGCGCTGCGCGCTGACCCACACCGCTTGGCCCTCGGCTTG 441
DB 1010 AGAGGGTGTTCCTTGTGGAGGCGCGTAGACGTGACGTCGCTTCCGCTGCGCG 951
QY 442 AGCGAGTTCAGCGGGCGAGCGGTGCGATGCGGATGCCCTCGACGTGGCGGCGAAC 499
DB 950 AGACCATGCAGCATCTTACGAGTTCTGCGCGGTGCGCGCGACGTTGTGCGGGAAC 893

RESULT 15
US-09-769-734-8/c
; Sequence 8, Application US/09769734
; Publication No. US2003014366A1
; GENERAL INFORMATION:
; APPLICANT: Ecopia Biosciences Inc.
; TITLE OF INVENTION: Genetic Locus for Evernimycin Biosynthesis
; FILE REFERENCE: PA 005-US
; CURRENT APPLICATION NUMBER: US/09/769, 734
; CURRENT FILING DATE: 2001-01-26
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 14252
; TYPE: DNA
; ORGANISM: M. carbonacea
; NAME/KEY: misc feature
; LOCATION: (459)..(1280)
; OTHER INFORMATION: ORF 5 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (2677)..(3747)
; OTHER INFORMATION: ORF 7 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (1280)..(2566)
; OTHER INFORMATION: ORF 6 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (3899)..(4774)
; OTHER INFORMATION: ORF 8 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (4893)..(5303)
; OTHER INFORMATION: ORF 9 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (5365)..(6306)
; OTHER INFORMATION: ORF 10 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (6350)..(7204)
; OTHER INFORMATION: ORF 11 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (7371)..(8198)
```

```
; OTHER INFORMATION: ORF 12 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (8304)..(9098)
; OTHER INFORMATION: ORF 13 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (9462)..(10493)
; OTHER INFORMATION: ORF 14 (positive strandedness)
; NAME/KEY: misc feature
; LOCATION: (10665)..(11384)
; OTHER INFORMATION: ORF 15 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (11387)..(12700)
; OTHER INFORMATION: ORF 16 (negative strandedness)
; NAME/KEY: misc feature
; LOCATION: (12971)..(14185)
; OTHER INFORMATION: ORF 17 (negative strandedness)
US-09-769-734-8

Query Match      18.5%; Score 92.4; DB 10; Length 14252;
Best Local Similarity 53.6%; Pred. No. 9.3e-14;
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 142 AGTGTACGCGGGGACGTACAGGATCCACTGTCCGCCAGCCCGGGGAACTCTCTCTCC 201
DB 2529 ACCCGCAGGCTGGGCGAGGAGATCAGCTTGTTCCGGCGCGCCACCGTGGCGCTCG 2470
QY 202 TTGCGCATGATCTCGTCGGCGTGTTCAGGCGAAGAGCAGCGGTAGTCCACCGCTCG 261
DB 2469 CGCGCCAGCATCTCGTCGTGGAAGTGCACCGCAGCACCAGTAGTGTTCGGCGCGG 2410
QY 262 GCGGTGAACCGCTCCGGGGTGGCGCACCGGATGTGCGTCCGGGGGTGAGCGGCGCTGC 321
DB 2409 GCCCGCAGTGGCTCGCTGATGATCTCGATGTCCGTACCGAGGTCCGGCGCGGACC 2350
QY 322 TTGCGCGCGTGTGTGCGACACCCAGGACAGGTCCGCCAGCATCGCGCAGATTC 381
DB 2349 TTGTCCGGTGTGCTTGTGAGGCGCGGTAGACGTGACAGGTGCTTCCGCTGCGGCG 2290
QY 382 GTACAGTGGCGCTCTTCGCGCTCCGCCGTACGCCACACCGCTTCCCTCTCGGCTTG 441
DB 2289 AGGAGGGTGTGCTTGTGAGGCGCGGTAGACGTGACAGGTGCTTCCGCTGCGGCG 2230
QY 442 AGCGAGTTCAGCGGGCGAGCGGTGCGTGGGATGCCCTCGACGTGGCGGCGAAC 499
DB 2229 AGACCATGCAGCATCTTACGAGTTCTGCGCGGTGCGCGCGACGTTGTGCGGGAAC 2172

RESULT 16
US-10-107-431-280/c
; Sequence 280, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Stafifa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORF
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 280
; LENGTH: 48221
; TYPE: DNA
; ORGANISM: Micromonospora carbonacea aurantiaca
; NAME/KEY: misc feature
; OTHER INFORMATION: "n" at position 106 is undetermined and represents a or g or t c
; OTHER INFORMATION: c
US-10-107-431-280

Query Match      18.5%; Score 92.4; DB 13; Length 48221;
```

Best Local Similarity 53.6%; Pred. No. 7.6e-14;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

QY 142 AGGTGTACGCGGGGACGTACAGGATCCACTGCGCGACGCCGCGAACTCTGCTCTCC 201  
Db 8244 ACCCGAGGCTGGGACGCGGGAAGATCAGCTTGGTTCCGGCGCCACCGTGGCGCTCG 8185

QY 202 TTGCGCATGATCTCGTCGCGGTGGTTCCAGGCGAAGAGAGCGGTAGTCCACCGCGTCG 261  
Db 8184 CGCGCCACCATCTCGTCGTGGAAGTCCACCGCAGCACAGTAGTGTCCGGCGCGCG 8125

QY 262 GCGGTGAACGCTCGCGGTGCGCACCGGATGTCGTCGCGGGGTGAGCGGCCCTGC 321  
Db 8124 GCGCGGAGTCCGCTCGCTGATGATCTCGATGTCGTCACGAGGTCCGGCGCCGACC 8065

QY 322 TTGCGCGGCTCGTGTGCGACACCGAGGAGACAGGTCCGACCGATCGCGAGAGTTC 381  
Db 8064 TTGTCGCGGTTCGCTCGCGCGGTACGCGATCAGCGTCCGTCGATCCCGAGTACTGC 8005

QY 382 GTACGGTGGCTCTTCGCGCTCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 441  
Db 8004 AGAGGGTGTTCGCTTGTGAGGCGCGGTGAGAGTGTGCGGTGCGGTGCGGTGCGGT 7945

QY 442 AGCGATTCAGCAGGCGGAGGAGTGGTGGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 499  
Db 7944 AGACCATGAGCATCTTGACGATTCGTCGCGGTGCGGTGCGGTGCGGTGCGGTGCG 7887

RESULT 17  
US-10-107-431-277/c  
; Sequence 277, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 277  
; LENGTH: 45055  
; TYPE: DNA  
; ORGANISM: Streptomyces mobaraensis  
US-10-107-431-277

Query Match 18.1%; Score 90.6; DB 13; Length 45055;  
Best Local Similarity 52.5%; Pred. No. 2.1e-13;  
Matches 198; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 122 CACCGCGGTGACGTCGTCACGTGTACGCGGGGACGTACAGGATCCACTGTCGCCAG 181  
Db 9974 CCGCGGACCGGTGACACCATCTTCCAGCGGGGACGCGGAGACAGGTTGCGCGCG 9915

QY 182 CCGCGGGAATCTCGTCTCTTCCGATGATCTCGTGGCGGTGTTCCAGGCGAAGACA 241  
Db 9914 CCGCGGGAATCTCGTCTCTTCCGATGATCTCGTGGCGGTGTTCCAGGCGAAGACA 9855

QY 242 GCGGTGATCCACCGGTGCGGTGAAACGCGTCCGGGTGCGCACCGGATGTGCTGC 301  
Db 9854 GGAACCGGTCCGCTCAGCGCCCGGGCTCTCTTCCGAGAGCATCGGATGTCCGTGC 9795

QY 302 CGGGGTTGAGCGCGCTGTTGCGCGGGTGTGTGCGACACCCAGGAGACAGTCCG 361  
Db 9794 CGGGGTTGAGCGCGCTGTTGCGCGGGTGTGTGCGACACCCAGGAGACAGTCCG 9735

QY 362 GACCGATGCGCGAGATGTCGTCAGCGTGGCGCTCTTCCGCGTGGCGCGTCCGACCA 421  
Db 9734 CGTCCAGCGCGATGTCGAGGATGACGTTGCGCTTGTGAGGCGCGTATCCCGAGG 9675

QY 422 CCGCTTGGCCCTGCGGCTTGACGAGTTCAGAGGGGAGAGGTCGCTGCGGATGCCCT 481  
Db 9674 TGGCGACGCCCTTGTGTCGCGAGGTGAGGAGTCCACGAGCCGCTGCGCGGATTCCT 9615

QY 482 CGACGTCGCGCGGGAAC 498  
Db 9614 CGGTCCGCGCGGGAAC 9598

RESULT 18  
US-10-107-431-118/c  
; Sequence 118, Application US/10107431  
; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 118  
; LENGTH: 1224  
; TYPE: DNA  
; ORGANISM: Streptomyces mobaraensis  
US-10-107-431-118

Query Match 17.6%; Score 87.8; DB 13; Length 1224;  
Best Local Similarity 52.6%; Pred. No. 1.8e-12;  
Matches 191; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 136 CTGTCACGTGTACGCGGGGACGTACAGGATCCACTGTCGCCAGCCCGCGGAACTCC 195  
Db 1223 CACACCATCTCCAGGCGGCGACGCGGAGACAGGTTGCGCGCGCGCCAGGAGTCG 1164

QY 196 TGTCTCTTCCGCGGTGTCGTCGCGGTGTCGCGGAGAGAGAGCGGTAGTCCACC 255  
Db 1163 CGTTCCCGGCGACCATCGCTCCCGTAGATCCACGCGGAGGACAGGAACTCCGCGC 1104

QY 256 GCGTCGCGGTGAACGCTCCGCGGTGCGCACCGGATGTCGCGCGGGGTGAGCCGG 315  
Db 1103 TCCAGCGCCCGGCTCTCTTCTCGAGACGATCGGATGTCGCGCGGTGAGGAGCAG 1044

QY 316 CCCTGCTTGGCGCGCTGTCGTCACACCCAGGAGACAGGTCGCGGACCGATGCGCGCAG 375  
Db 1043 CCGAACCTTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 984

QY 376 AGTTCGTCAGGTGGCGCTTTCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 435  
Db 983 TACTGGAGGATGAGTTCGCTTGGTGGAGGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 924

QY 436 GCGTTGAGCGAGTTCAGCAGGCGGAGCAGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 495  
Db 923 TCCCGCGAGGTGTCGAGGAGTCCACGAGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 864

QY 496 AAC 498  
Db 863 AAC 861

RESULT 19  
US-10-229-148B-1/c  
; Sequence 1, Application US/10229148B  
; Publication No. US20040091975A1  
; GENERAL INFORMATION:  
; APPLICANT: Meiji Seika Kaisha, Ltd.  
; TITLE OF INVENTION: Midcamycin biosynthetic genes  
; FILE REFERENCE: 138451 US  
; CURRENT APPLICATION NUMBER: US/10/229,148B  
; CURRENT FILING DATE: 2002-08-28

```

; PRIOR APPLICATION NUMBER: 210516/2002
; PRIOR FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 1
; LENGTH: 84428
; TYPE: DNA
; ORGANISM: Streptomyces mycarofaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((11)..(673))
; OTHER INFORMATION: ORF42 (fragment)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((1168)..(2202))
; OTHER INFORMATION: ORF41
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((2220)..(3215))
; OTHER INFORMATION: ORF40
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((3237)..(4691))
; OTHER INFORMATION: ORF39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((4695)..(5948))
; OTHER INFORMATION: ORF38
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6048)..(6629))
; OTHER INFORMATION: ORF37
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((6653)..(7945))
; OTHER INFORMATION: ORF36
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8149)..(9015)
; OTHER INFORMATION: ORF35
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9012)..(9335)
; OTHER INFORMATION: ORF34
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9328)..(10458)
; OTHER INFORMATION: ORF33
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10521)..(11603)
; OTHER INFORMATION: ORF32
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11729)..(12961)
; OTHER INFORMATION: ORF31
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((13016)..(14044))
; OTHER INFORMATION: ORF30
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((14074)..(15096))
; OTHER INFORMATION: ORF29
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15643)..(17466)
; OTHER INFORMATION: ORF28
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((17522)..(18895))
; OTHER INFORMATION: ORF27
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19063)..(20229)
; OTHER INFORMATION: ORF26
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((20307)..(21743))
; OTHER INFORMATION: ORF25
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((21733)..(22527))
; OTHER INFORMATION: ORF24
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((22534)..(23571))
; OTHER INFORMATION: ORF23
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((23555)..(24463))
; OTHER INFORMATION: ORF22
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((24460)..(25650))
; OTHER INFORMATION: ORF21
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((25647)..(26105))
; OTHER INFORMATION: ORF20
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Complement((26180)..(27391))
; OTHER INFORMATION: ORF19
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (27937)..(28983)
; OTHER INFORMATION: ORF18
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (29244)..(42779)
; OTHER INFORMATION: ORF1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42823)..(48657)
; OTHER INFORMATION: ORF2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (48712)..(59802)
; OTHER INFORMATION: ORF3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (59850)..(64556)
; OTHER INFORMATION: ORF4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (64687)..(70365)
; OTHER INFORMATION: ORF5
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70365)..(71078)
; OTHER INFORMATION: ORF6
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71113)..(72360)
; OTHER INFORMATION: ORF7
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72400)..(73665)
; OTHER INFORMATION: ORF8
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (73694)..(75043)
; OTHER INFORMATION: ORF9
; FEATURE:
; NAME/KEY: CDS

```

LOCATION: Complement((75899)..(76570))  
OTHER INFORMATION: ORF10  
FEATURE:  
NAME/KEY: CDS  
LOCATION: Complement((76602)..(77765))  
OTHER INFORMATION: ORF11  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (78039)..(79313)  
OTHER INFORMATION: ORF12  
FEATURE:  
NAME/KEY: CDS  
LOCATION: Complement((79391)..(81052))  
OTHER INFORMATION: ORF13  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (81541)..(82356)  
OTHER INFORMATION: ORF14  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (82760)..(83362)  
OTHER INFORMATION: ORF15  
FEATURE:  
NAME/KEY: CDS  
LOCATION: Complement((83495)..(84142))  
OTHER INFORMATION: ORF16  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (84329)..(84428)  
OTHER INFORMATION: ORF17 (fragment)  
US-10-229-148B-1

Query Match 16.8%; Score 84.2; DB 17; Length 84428;  
Best Local Similarity 51.2%; Pred. No. 7e-12;  
Matches 197; Conservative 0; Mismatches 188; Indels 0; Gaps 0;  
QY 114 CGGGCGGACACGGGGGTGACGTCAGCTGACGTGACGGGGGAGGTACAGATCCACTG 173  
Db 12982 CGGGCGGGCGGGCGGCGCTGTGCACAGACGCTCAGCTCGGGCAGCGGAAGAGAGTTT 12923  
QY 174 TCGCGCAGCCCGCGGAATCTCTGCTTCCGCAATGATCTGCTGGGGTGTTCAGGCG 233  
Db 12922 GCGCGGGCTCCCGGTAGCGCGCTCCCGCTCAGGAGCGCTCGCATAGATCCACGG 12863  
QY 234 GAAGAGACGCGTGTGTCACCGCTCGGGCGGTGAACGCGCTCGGGGGTGGCGACCGGAT 293  
Db 12862 CAGCACCAGCAGTTGGTCGGGCTTGAGGAGCTTGCGCTCTCTCGGACACGATCGGGAT 12803  
QY 294 GTGCGTGGCGGGGTGAGCGGCGCTGTGTCGGCGGCTGCTGCTGACCTCGCGATCGGGCAG 353  
Db 12802 CGCGATCCGGGGTGAAGCGCGCTTCTCTGCTGACCTCGCGATCGGGCAG 12743  
QY 354 CAGGTCGACCGATGCGCGAGAGTTGCTACGCTGGCGCTCTTCCGCGTGGCGCGGTA 413  
Db 12742 GTCCGCTCGGTGAGACCGCAGTACTGGAGGATCAGTTGCCCTTGGTGGAGGCGCCGTA 12683  
QY 414 CGCACACCGCTTCCCTCGCGCTTGAGCGAGTTACAGCGGGGAGCAGTGGTGGTGG 473  
Db 12682 TCACCGGTGAGACGCGCGCTCGGGGAGCGGTGAGGAGTGCATCAGGGGCTCGCG 12623  
QY 474 GATGCGCTCAGCTCGGGCGGAAC 498  
Db 12622 CTGGTCTGACACGGCGCGGAAC 12598

RESULT 20  
US-10-084-846A-1/c  
Sequence 1, Application US/10084846A  
Publication No. US2004006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MUHLENWEG, AGNES  
APPLICANT: TREFZER, AXEL

APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PCT/EP01/09815  
PRIORITY APPLICATION NUMBER: DE 101 09 166.4  
PRIORITY FILING DATE: 2001-08-24  
PRIORITY APPLICATION NUMBER: DE 101 09 166.4  
PRIORITY FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 1  
LENGTH: 59816  
TYPE: DNA  
ORGANISM: Streptomyces viridochromogenes  
US-10-084-846A-1  
Query Match 16.2%; Score 81; DB 16; Length 59816;  
Best Local Similarity 50.9%; Pred. No. 4.5e-11;  
Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;  
QY 122 CACCGCGGGTCACTGTCAGCTGTACCGGGGGAGGTACAGGATCCACTGTCCGCCAG 181  
Db 15020 CTCGCCACCCGATCACAACTTCAGCGTGGCGACGCGAAGACAGGCTGCTCCCG 14961  
QY 182 CCGCGCGGAATCTCTCTCTTCGCCATGATCTCGTCGGCTGTTTCCAGGCGAAGACA 241  
Db 14960 AGGCCAGGAAGTCGCTTCCCGGGCAGCATCGGTCCTCCGATCCACGCGAGGACCA 14901  
QY 242 GCGGTAGTCCACCGGTGCGGGGTGAACGCTCGGGGTGCGACCGGGATGCGGTGC 301  
Db 14900 GGAATGTCGGGCTCAAGCGCGGGCTCTCTCTCGGAGACGATCGGGATGTCGTGC 14841  
QY 302 CGGGGTGAGCGCGGCTCTTGGCGGGTGTGTCGCGACACCCAGGAGACCGAGTCCG 361  
Db 14840 CGGCGGTGAGCAGCGCGAATCTCTCTTCACTCGCGATGACAGGCGAGGCTCT 14781  
QY 362 GACCGATGCGCGAGAGTTCGTCACGCTGCGCTCTTCCCGTGGCGGCTAGCCACCA 421  
Db 14780 CGTCCAGCGCGAGTACTGCGATGACGTTGCCCTTCGTGAGGCGCGCTAGCCAGG 14721  
QY 422 CCGCTTGCCTCGGCTTTCAGCGAGTTTCAGAGGCGGAGCAGGTCGTCGCGATGCCCT 481  
Db 14720 TGTGACGCGCTTGTCCCGGCGGCTGAGGATTCAGGATCGTCCAGGACTCCT 14661  
QY 482 CGACGTGCGCGGCAAC 498  
Db 14660 CGGTCCCGCGCGCAAC 14644

RESULT 21  
US-10-084-846A-2  
Sequence 2, Application US/10084846A  
Publication No. US2004006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MUHLENWEG, AGNES  
APPLICANT: TREFZER, AXEL  
APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PCT/EP01/09815  
PRIORITY APPLICATION NUMBER: DE 101 09 166.4  
PRIORITY FILING DATE: 2001-08-24  
PRIORITY APPLICATION NUMBER: DE 101 09 166.4  
PRIORITY FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 2  
LENGTH: 59816  
TYPE: DNA  
ORGANISM: Streptomyces viridochromogenes





```
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 2058
/ LENGTH: 1671
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1671)
US-10-156-761-2058

Query Match      11.4%; Score 57; DB 15; Length 1671;
Best Local Similarity 50.1%; Pred. No. 5.8e-05;
Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

QY 141 CACGRTAGCGGGGAGCGTACAGGATCCACTGTCGCGAGCCCGCGGAACTCCTGTCTC 200
Db 1563 CAGGTGGAGAGCGCGGGCACAGCGCGTCCACGAAACCGTCGACGTCCAGAGACGCTCCAG 1504
QY 201 CTTGCCATGATCTCGTCGGGTGTTCCAGCGGAAGAGCGGCGGTAGTCCACCGGCTC 260
Db 1503 GGTCCGGATCAGTCGGCGCGGTGCGCGGTGCTGTAATCCCGCAGGGGTGACGACAG 1444
QY 261 GGGCGTGAACCGGTCCGGGGTCCGACCCGCGATGTGCGTCCGGGGTGAAGCCGCCCTG 320
Db 1443 GCGCGTGAAGCGCGCGCGTCTCTCCGGGACGAA-----GGGAGCAGCAGGACGTG 1390
QY 321 CTTGCCCGCGTCTGTCGCGGTGTTCCAGCGGAAGAGCGGCGGTAGTCCACCGGCTC 260
Db 1503 GGTCCGGATCAGTCGGCGCGGTGCGCGGTGCTGTAATCCCGCAGGGGTGACGACAG 1444
QY 261 GGGCGTGAACCGGTCCGGGGTCCGACCCGCGATGTGCGTCCGGGGTGAAGCCGCCCTG 320
Db 1443 GCGCGTGAAGCGCGCGCGTCTCTCCGGGACGAA-----GGGAGCAGCAGGACGTG 1390
QY 321 CTTGCCCGCGTCTGTCGCGGTGTTCCAGCGGAAGAGCGGCGGTAGTCCACCGGAGTT 380
Db 1389 CGAGGCCAGCTCTGTCGGCGCGCGCGCAGACCCGCGGGCGGGCGGGCGCCACCG 1330
QY 381 COTCAGCGTGGCGCTCTTCGCGCGTCCGCGCGTACGCCACCCCGCTTGCCTTCGGCTT 440
Db 1329 GCGGCGGTGCGCGGCTCTCTCCAGGGCGCGCGCAGCCCTCCGCGGAGTGCACGGCG 1270
QY 441 GAGCGAGTTCAGCAGCGCGCAGCGTCCGTGCGGATGCGTCCGGATGCGCTCG 483
Db 1269 GCTGACCCCGAGCGTGAGCGCCCGTCTGCTTCAGGCGCCGCG 1227

RESULT 25
US-10-156-761-1/c
/ Sequence 1, Application US/10156761
/ Publication No. US20030119018A1
/ GENERAL INFORMATION:
/ APPLICANT: OMURA, SATOSHI
/ APPLICANT: IKEDA, HARUO
/ APPLICANT: ISHIKAWA, JUN
/ APPLICANT: HORIKAWA, HIROSHI
/ APPLICANT: SHIBA, TADAYOSHI
/ APPLICANT: SAKAKI, YOSHIYUKI
/ APPLICANT: HATTORI, MASAHIRO
/ TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
/ FILE REFERENCE: 249-262
/ CURRENT APPLICATION NUMBER: US/10/156,761
/ PRIOR FILING DATE: 2002-05-29
/ PRIOR APPLICATION NUMBER: JP 2001-204089
/ PRIOR FILING DATE: 2001-05-30
/ PRIOR APPLICATION NUMBER: JP 2001-272697
/ PRIOR FILING DATE: 2001-08-02
/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 1
/ LENGTH: 9025608
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (4187715)
/ OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      11.4%; Score 57; DB 15; Length 9025608;
Best Local Similarity 50.1%; Pred. No. 1.4e-05;
Matches 172; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

/ NUMBER OF SEQ ID NOS: 15109
/ SEQ ID NO 2058
/ LENGTH: 1671
/ TYPE: DNA
/ ORGANISM: Streptomyces avermitilis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1671)
US-10-156-761-2058

Query Match      11.2%; Score 56.2; DB 15; Length 2982;
Best Local Similarity 47.5%; Pred. No. 8.3e-05;
Matches 233; Conservative 0; Mismatches 253; Indels 5; Gaps 2;

QY 7 CGACCGGTGTCGCGGAAACAGAGTCAGATCGAGATACGCGGAGGAAACACCCCGGTAGTCCG 66
Db 1087 CGCGCGGTCTCTTGAACCTCAGGTACGGGATGAGCGTGAACCGCATGTCTGACCG 1028
QY 67 GGTAGCGGTGGGCGGAGCGGTACGCGCTTCGACGTCACGGGCGG--CGCGACAC 124
Db 1027 GGAAGTCCAGTGTGGGCGCGGTAGCGGCCACCGCGCTGACTGGCGGAAACGGGTCT 968
QY 125 CGCGCGGTGAGCTCGTCACTGTACGCGGGGAGCGTACAGGATCCATGTCCTCCAGCCCC 184
Db 967 CGCGCGGCGAGTTGGCGGCGACCCAGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 908
QY 185 GCGGGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 244
```

Db 907 CTTCTGCTACTGTCGAGCAGATGTCGCTTCTCGCGGTCTTCGCGGAGAACTCGACGA 848  
QY 245 CTTAGTCCACCGGTTCGCGGTGAAACGCTCCGGGTGCGCACCGGGATGTGCGTCCCG 304  
Db 847 CCAGGTCAACTGTCGCGGCCCCCAGCGCGCGAGCAGCACCGGCTCGGCGGACCG 788  
QY 305 GGTGAGCGCGCTGCTTGGCGGCTGCTGTCGACACCCAGGAGACCAAGTCCGGAC 364  
Db 787 G---CAGTCCGCGTGGAGCGCTTGGCGATCAGCCAGACCACTTCTCAGCGGTGAGCG 731  
QY 365 CGATCGCGCAGAAAGTTCGTCACGCTGCGCTCTTTCGCGCTGCGCGCTAGCCACCAACC 424  
Db 730 GGGTCCCTTGGAGCCCTCCAGGCTCTCGTGGACCTCGATCGCGACCTGCTCCACTCG 671  
QY 425 GCTTGCCTCGCTTGGAGCTTACAGAGCGGAGCAGGTGCTGCGGATCCCTTCGA 484  
Db 670 CGAGGACGAGGGGACAGGTACTGTCGCGGACCCCGGTGCGGTGGAGACCGCCTGGT 611  
QY 485 CGTCGCGCGCG 495  
Db 610 CGGCGCGCGCG 600

RESULT 27  
US-10-156-761-1  
; Sequence 1, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 1  
; LENGTH: 9025608  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4187715)  
; OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Query Match 11.2%; Score 56.2; DB 15; Length 9025608;  
Best Local Similarity 47.5%; Pred. No. 2.2e-05;  
Matches 233; Conservative 0; Mismatches 253; Indels 5; Gaps 2;

QY 7 CGACCGTTCGCGGAAACAGTCGAGATACGCGGAGAGGAAACACCCCGGTAGTCGG 66  
Db 5495898 CGGCGCGGTCTTGAACCTCAGGTACGGGATGAGCGTGAACCGCATGTCGATCTCGACCG 5495957  
QY 67 GGTAGACGCTGGCGCGGAAGCGGTACGCGCTTCGACGGTTCAGCGGCGG--CGCGACAC 124  
Db 5495958 GGAAGTCCAGTGTGGCGGCGGTAGCGCGCCACCGCGTGCAGTGGCGAAGCGGTCT 5496017  
QY 125 CGGCGGTTCAGTCTGTCAGTGTACGCGGGGAGGTACAGGATCCACTGTCCCGCACGCC 184  
Db 5496018 CGGCGGCGCACTTGGCGGCGACCAAGGCTGTGGTGTAGTGTCTGTCTGCTCGCGGTGA 5496077  
QY 185 GCGGAACTCTGCTCTTCTGCCATGATCTCGTGGCGGTGTTCCAGGCGAAGACGACG 244

Db 5496078 CTTCTGCTACTGTCGAGCAGATGTCGCTTCTTCGCGGTCTTCGCGGAGAACTCGACGA 5496137  
QY 245 CTTAGTCCACCGGTTCGCGGTGAAACGCTCCGGGTGCGCACCGGGATGTGCGTCCCG 304  
Db 5496138 CCAGGTCAACTGTCGCGGCCCCCAGCGCGCGAGCAGCACCGGCTCGGCGGACCG 5496197  
QY 305 GGTGAGCGCGCTGCTTGGCGGCTGCTGTCGACACCCAGGAGACCAAGTCCGGAC 364  
Db 5496198 G---CAGTCCGCGTGGAGCGCTTGGCGATCAGCCAGACCACTTCTCAGCGGTGAGCG 5496254  
QY 365 CGATCGCGCAGAAAGTTCGTCACGCTGCGCTCTTTCGCGCTGCGCGCTAGCCACCAACC 424  
Db 5496255 GGGTCCCTTGGAGCCCTCCAGGCTCTGTCGACCTCGATCGCGACCTGCTCCACTCG 5496314  
QY 425 GCTTGCCTCGCTTGGAGCTTACAGAGCGGAGCAGGTGCTGCGGATCCCTTCGA 484  
Db 5496315 CGAGGACGAGGGGACAGGTACTGTCGCGGACCCCGGTGCGGTGGAGACCGCCTGGT 5496374  
QY 485 CGTCGCGCGCG 495  
Db 5496375 CGGCGCGCGCG 5496385

RESULT 28  
US-10-282-122A-15092/c  
; Sequence 15092, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malore, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15092  
; LENGTH: 837  
; TYPE: DNA  
; ORGANISM: Bordetella pertussis  
US-10-282-122A-15092

Query Match 11.0%; Score 54.8; DB 13; Length 837;  
Best Local Similarity 46.2%; Pred. No. 0.00022;  
Matches 182; Conservative 0; Mismatches 212; Indels 0; Gaps 0;

QY 106 TCAGCGGCGGCGGACACCGCGGGTTCAGTCTCAGTGTACGCGGGGAGGTACAGG 165  
Db |||||  
QY 539 TCAGTGGCTGGCTGGCGACGATGCCGTAATTGGCGGTGTCTTCGCGCGGACGCTCCTGT 480  
Db |||||  
QY 166 ATCCACTGTCCGCCAGCGCGCGGAACTCCCTGCTCCCTTCGCCATGATCTGTTCGGCGTGG 225  
Db |||||  
QY 479 ACGCCAGCAGCTGCCCTTGGCGGCGCACCGCAGCTGCATCAGTCTGTAGCACCGCGC 420  
Db |||||  
QY 226 TTCAGGCGAAGAGCAGCGCGTAGTTCACCCCGTTCGGCGGTGAACGCGTTCGGGGTCCGC 285  
Db |||||  
QY 419 GTGTCCGCGTGCATCAGGTCTATCGCCAGCAGCACCGCGAAGGGCTGTTGCCCTACCGCC 360  
Db |||||  
QY 286 ACCGGGATGTGCTGCCGCGGGTTCAGCGCGGCTTTCGGCGCGCTGTGTCACAC 345  
Db |||||  
QY 359 GCGCGCGCGTTCAGCACGCGATGTCCAGCGCCAGCGGGCGGATTTGGCGGATGTACAGG 300  
Db |||||  
QY 346 CAGGAGACAGGTTCGGACCGATGCCGAGAACTTCGTCAGGTGGCGCTTTCGCGCTC 405  
Db |||||  
QY 299 CAGTTACGTGCGCGGAGGATGTCCGGGACGAGCGCGCAGTTCGTGCTTCCCTTC 240  
Db |||||  
QY 406 GCGCGGTACGCACACCGCTTGCCTCGGCTTCAGCGGTCAGCAGGCGGAGCAGG 465  
Db |||||  
QY 239 GCTTCCAGTTCGTTTCAGCTTCGGGGCGGATCGAATGCTTCGATGGCGCGCTTG 180  
Db |||||  
QY 466 TCGGTTCGATGCCCTTCAGCTTCGGCGCGCAACC 499  
Db |||||  
QY 179 TTGCGCGCGTGACGAGATGAGTTCGTGTATCC 146  
Db |||||

RESULT 29  
US-10-425-114-24834/c  
; Sequence 24834, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; NUMBER OF SEQ ID NOS: 2003-04-28  
; SEQ ID NO 24834  
; LENGTH: 1368  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3689-204-A2\_FLI  
US-10-425-114-24834

Query Match 10.9%; Score 54.4; DB 13; Length 1368;  
Best Local Similarity 49.0%; Pred. No. 0.00026;  
Matches 145; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 200 CTTTGCCATGATCTTCGTGGCGGTTCAGCGGAAAGAGCAGCGCGTAGTCCACCGCGT 259  
Db |||||  
QY 1132 CCGAGCGGTGAGCAGGCGAGAGTCTTCACCCGAGCTTCAGCGCGCGCTTCCTCCA 1073  
Db |||||  
QY 260 CCGGGCTGAACGCTCCGGGTGGCGACCGGATGTGGTGCCTGGGGGTGAGCGGCGCT 319  
Db |||||  
QY 1072 CCAGCGTGTACCGGTTCGGAGGGGAACTGCCAGCAGGAGGACGCTTCGCGCGCGGC 1013  
Db |||||  
QY 320 GCTTGCGCGCGCTGCTGTCGACACCCAGGAGACAGGTTCGCGACCGATGCGCGAGAA 379  
Db |||||  
QY 1012 TGC CGCGCAGCAGGCGGCGCGGAAACCGCGACTGCGCGCAGTTCCTCGCGCCAGCTGCGGA 953  
Db |||||

Query Match 10.8%; Score 53.8; DB 15; Length 1734;  
Best Local Similarity 49.3%; Pred. No. 0.00035;  
Matches 169; Conservative 0; Mismatches 172; Indels 2; Gaps 1;

QY 100 CGAGCGTTCAGCGCGCGGCGGACACCGCGGGTTCAGTCTCAGCTGTACGCGGGGAGC 159  
Db |||||  
QY 869 CGCGGGAGCTTCGCGGCTACGCGGACCTGAGCGCGATCGCTGCGGCTGGCGCGC 928  
Db |||||  
QY 160 TACAGATTCATGTTCGCGGAGCGCGGAACTCTCTGCTCTCTGCTGCTGCTGCTGCTG 219  
Db |||||  
QY 929 TCGAGGAGCGCGCGCGCTCACCCAGCTGACCCGCAAGTACGCGCCAGGACATCGAG 988  
Db |||||  
QY 220 GCGTGGTTCAGGCGAAGAGCAGCGGTAGTCCACCGCTCGGCGGTGAACGCGTCCGG 279  
Db |||||  
QY 989 GCGTGTGACTGGCGCGGAGAGAGCGCGCGGCTCACCGAATCTCGAGCGGAGAGC 1048  
Db |||||  
QY 280 GTCCGACCGGATGTGCTGCGCGGGGTGAGCGGCGCTCTGCTTGGCGCGGCTGCTGCTG 339  
Db |||||  
QY 1049 AGCGGATTCGCGGAGCTGACCGCGGAGCGCGCTGCGGACCGAATCTGGCGCGGCTCG 1108  
Db |||||  
QY 340 CACACCCAGGAGACAGGTTCGCGGATGCGCGGAGAA--GTTCTGCTACGCTGGCGCTCT 397  
Db |||||  
QY 1109 CAGAGCGGTTCAGGAGCGCGGAGCGGAGCGCGGAGCGCGGAGCGGCTTCGCGCGCGG 1168  
Db |||||  
QY 398 TCSCCGCTCGCGCGTACGCCACCCCGCTTCGCCCTTCGCCCTTCGCCCTTCGCCCTTC 440  
Db |||||  
QY 1169 CGGAGCTGGCGTCTGCTCGCATGCCGATGCCGACGCGCGCTGCTGCTT 1211  
Db |||||

RESULT 31  
US-10-156-761-2271/c  
; Sequence 2271, Application US/10156761  
; Publication No. US20030119018A1

```

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2271
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1137)
US-10-156-761-2271

Query Match      10.7%; Score 53.6; DB 15; Length 1137;
Best Local Similarity 49.3%; Pred. No. 0.00042;
Matches 140; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 215 CGTCGGCGTGTCCAGCGAAGACGCGGTAGTCCACCGCGTGGGGGTGAACGGCT 274
Db 1030 CGTGGCGGAGCCACCGATGAATCCAGCGCGCGAGCACTCCAGGACCGGGCGCGA 971

QY 275 CCGGGGTGCGACCGGATGTGCGTCCGGGGGTGAGCCGCGCTTGTGGCGGGTGG 334
Db 970 TCCGCGCAGCGCTTCTGACAGTGTAGCGGCGCAGGAGCGCGGCGACCGACTCT 911

QY 335 TGTGCGACACCGAGGACAGGTCCGACCGATGCCGAGAGTTGCTACGGTGGCGC 394
Db 910 CGTCGCGCGCTCCGTGCCGATCGCCCGTGCAGACACCTCGAGAGCGCCAGCGCGCT 851

QY 395 TCTTCGCGCTGGCGCGTACGCCACCGCTTGGCCCTCGGCGCTTGGAGAGTTGAGCA 454
Db 850 CCATCTCGACCGGAGCTCGGCGCGCTTCTGAGAGCTGCGCGCTCCCGCTCCAGCACT 791

QY 455 GCGGAGCAGGTGCGTGGGATGCCCTCGACGTCGCGCGGAAC 498
Db 790 GCTCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 747

```

## RESULT 32

```

US-10-156-761-6854/c
; Sequence 6854, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```

```

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6854

```

```

; LENGTH: 540
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(540)
US-10-156-761-6854

```

```

Query Match      10.7%; Score 53.4; DB 15; Length 540;
Best Local Similarity 47.1%; Pred. No. 0.00053;
Matches 235; Conservative 0; Mismatches 256; Indels 8; Gaps 2;

QY 5 CCGGACCGTGTCCCGGAACACGAGTCAGATACGCGGAGAGAAACACCCCGGTAGTC 64
Db 510 CCTGGCCTCGTCGCGAGCGTGGGCTCTCCCGCGCTCGGCTGGGTGGCATGGAGCGGCG 451

QY 65 CCGGTAGACGGTGGCGGGAAGCGGTACGCGCTTCGACGCGTACGCGGCGGCGGACAC 124
Db 450 CGCGGGACCGTCCGACACCGTGGGATGCCGGTGTGACGCTCACCGTACGAGCGGCGT 391

QY 125 CGCGGGTACGTCGTACGTCAGTGTACGCGG-----GGACGTACAGGATCCACTGTCCGCC 179
Db 390 CGGCTCGGTCCGCGCTCGGCGTCCGCGGCTCGGCGGCTCGGCGGCGGCGGCGGCGGCGTA 331

QY 180 AGCCCGCGGAACCTCTCTCTCTCGCCATGATCTCTCGCGCTGTTTCAGGCGAAGAG 239
Db 330 ACGCGGTCGACCGCGCTCTGTGTGACTCCGAGCGGAGGCCACCGGCTCCACGAGAA 271

QY 240 CAGCGGTAGTCCACCGCGTGGCGGTGAACGCGTCCGGGTGGCGACCGGAGTGTGGT 299
Db 270 GCCCAGGAGCGGTGGAAGTCCACGCGCGCGCTGACAGGCTCTCGACACTGTGCGCGAG 211

QY 300 GCGGGGTGAGCGCGCGCTCTTTGGCGGCGTGTGTCGACACCGGAGGACGAGTC 359
Db 210 TTCTGGCGAGGCGGACGCTCGGAGCGGAGCGCGTCCGTAGACGAGAACCGCGTGA 151

QY 360 CGGACCGATGCCGAGAGTTGCTACCGTGGCGCTCTTCGCCGT---CGCGCGGTAGCG 416
Db 150 CGGCGGAGCGGCGGTGGCGGTAGAGCTTCCCAACTGGCGGTGAGCGTGCAGTGC 91

QY 417 CACCACCGCTTCCCTCGGCTTTGAGGAGTTGAGGAGTTGAGGAGGCGGAGGAGTGGTGGAT 476
Db 90 GTCCACTCTCCCGGAGCCCGCTCGATGTCCCGCACCGCAAGTGCAGGCTGGCGCGGCG 31

QY 477 GCCCTGACGTCGCGGCGG 495
Db 30 CTGGCGCTCGTGGGTGG 12

```

## RESULT 33

```

US-10-156-761-6184/c
; Sequence 6184, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

```

```

; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6184
; LENGTH: 894
; TYPE: DNA

```



Db 1313 TAGCCAGGCGCGTCCGCTCCCGGCGCGCTCCCGGCGAGACCGGAGACACCCCGTGGACG 1254  
Qy 96 CTTTGAACGGTCAAGCGGCGCGGCGAGACACCGCGCGGTCAAGTCTGTCAGTGTACCGCGG 155  
Db 1253 GCGGTCGCGG-CGACGGCGCGGAGACAGGACCGCGCGGTCTGTCAGCGGCGTCCCGAC 1195  
Qy 156 GAGTACAGATCCACTGTCCCGCAGCCCGCGGCGGACTCCTCTTCGCCATGATCTC 215  
Db 1194 CAGCCCGAGGCGAGCGGTGTCAGTGTCCCGTGTGACCGCGGTACCGAGCGTCTGGCATG 1135  
Qy 216 GTCCGCGTGTTCAGAGCGGAGAGAGAGCGGTAGTCCACCGCGTGTGGCGGAGAGCGGTC 275  
Db 1134 GCGGCGCGGTCCCGCGCGAGCGGTGCACATCCGCGCACACCGCGTACCGCAGCGCGGT 1075  
Qy 276 CGGCGTCCGACCGGAGATGTGTCGCGCGGCGGTGACCGCGCGTCTGTCGCGGCGTGT 335  
Db 1074 CGCGTTCGCGGCGCGGTGTGTCGCGAGCGGCGGCGACCGCGTGCAGCGCGGTCAAGCGAC 1015  
Qy 336 GTCCGACACCCAGGAGACAGGTCCCGACCGATGCCGAGAGTTCGTCAAGTTCGTCGCGT 395  
Db 1014 CGCGCGCACACCGCGCGCGCGTCCCGTCCGTCGACGACCGCGGTGACCGTCCGCGTACGCC 955  
Qy 396 CTTCCGCGTCCGCGGTGACGACACCGCGTTCGCGCGTTCGCGCGTTCGAGGAGTTCAGCAG 455  
Db 954 CTTGACCGTCCCGGCG 895  
Qy 456 GCGGACGAGTCCGTCGCGATGCCCTCGACGTCGCGCGCGGAA 497  
Db 894 GACCCCGTGACCGTCCCGTGTGAGTGGCGGCGCGCGCGG 853

## RESULT 36

US-10-239-079-1  
; Sequence 1, Application US/10239079  
; Publication No. US2003014846A1  
; GENERAL INFORMATION:  
; APPLICANT: Merck Patent GmbH  
; TITLE OF INVENTION: ANIC-BP-1-ligand  
; FILE REFERENCE: ANIC-BP-1-ligand  
; CURRENT APPLICATION NUMBER: US/10/239,079  
; CURRENT FILING DATE: 2002-09-19  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 2700  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (363)..(2432)  
; FEATURE:  
; OTHER INFORMATION: Description of ANIC-BP-1 protein ligand

US-10-239-079-1

Query Match 10.6%; Score 53.2; DB 15; Length 2700;  
Best Local Similarity 49.6%; Pred. No. 0.00045;  
Matches 136; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
Qy 218 CGCGGTGTTCAGGCGAGAGAGAGCGGTAGTCCACCGCGTCCGCGGTGAACGCGTCCG 277  
Db 326 CGCGCGGCGCGCGCGCGGTGCGGTGCGTCCGCGCGAGATGGGCGCGGTGCGCACCGGCCA 385  
Qy 278 GGGTGGCGACCGCGAGTGTGTCGCGCGGCGTGAAGCGCGCGTCTTGGCGCGGTGCTGT 337  
Db 386 TGGTGGGTGCGGCGTGGCG 445  
Qy 338 CGCACACCCAGGAGACAGGTCCGAGCCGATGCCGAGAGTTCGTCAAGTTCGCGCGTCT 397  
Db 446 AGACCGCGTTCGAGCGCGCGTCCCGGAGTGTGCGCGCGCGTGTCCGCGGTGAAGCAGG 505  
Qy 398 TCSCCGTCCGCGGTAGCGCACCGCGTTCGCGCGTTCGCGCGTTCGCGCGGTTCAGCGAG 457  
Db 506 CGGCG 565

Qy 458 CGAGCAGGTCCGTGCGGATGCCCTCCGACGTCGCG 491  
Db 566 CTTGCTGTCTGCGCGCGCGTGGCGCGCGAGGTAGCC 599

## RESULT 37

US-10-259-165-419/c  
; Sequence 419, Application US/10259165  
; Publication No. US20030135888A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; APPLICANT: Chang, Hur-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiyaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricketts, Darrell  
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING  
; FILE REFERENCE: 70030-NP  
; CURRENT APPLICATION NUMBER: US/10/259,165  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/368,327  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
; SEQ ID NO 419  
; LENGTH: 915  
; TYPE: DNA  
; ORGANISM: Oryza sativa  
; US-10-259-165-419

Query Match 10.6%; Score 52.8; DB 15; Length 915;  
Best Local Similarity 49.3%; Pred. No. 0.00068;  
Matches 138; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

Qy 210 GATCTCGTCCGCGGTTCAGGCGAGAGAGAGCGGTAGTCCACCGCGTCCGCGGTGAA 269  
Db 306 GATCCCGTACTCGTCCCGCGCGGTTCGATCGCGCGCGCGCGCGCGCGCGCGCGCG 247  
Qy 270 CGCGTCCGCGGTGCGCACCGCGGTGCGTCCGCGCGGTGCGCGCGGTGCGCGCGGT 329  
Db 246 GCAGTCTTCGATCGGAGCGGTGCGGAGCGAGTCCGCGGTTCGATCGCGGTTCGCGG 187  
Qy 330 CGTCTGTTCGCACACCGCGAGACCGGTCCGCGCGGTGCGCGCGGTGCGCGCGGT 389  
Db 186 GTGCGTCCGTCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 127  
Qy 390 GCGCTCTTCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 449  
Db 126 GATGTCACCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCG 67  
Qy 450 CAGCAGCGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 489  
Db 66 CATCACGTCGATCAGGTACCGCGCGGTGCTACTTGAGCTTG 27

## RESULT 38

US-10-259-165-79/c  
; Sequence 79, Application US/10259165  
; Publication No. US20030135888A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun

```

; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 79
; LENGTH: 918
; TYPE: DNA
; ORGANISM: Oryza sativa
; US-10-259-165-79

Query Match 10.6%; Score 52.8; DB 15; Length 918;
Best Local Similarity 49.3%; Pred. No. 0.00068;
Matches 138; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 210 GATCTCGTCGGCGTGGTTCCAGGCGAAGAGCAGCGCGTAGTCCACCGCGTGGCGGTGAA 269
DB 306 GATCCGCTACTCGTCCACCGCGACTTGATCGCACCGCAGTCGTGCGCGGACACAGATGTA 247
QY 270 CGCGTCGGGTGCGCACCGGATGTGCTGCGCGGGGTGAGCCGCCCTGCTTGCGCGG 329
DB 246 GCAGTCTCGATCGCGACGTGGAGCACAGTCCCGGGTGTGATCGCTCGGTGTTGGGCGA 187
QY 330 CGTCGTGCGCACACCCAGGAGACAGAGTCCCGGACCGATGCGCGAGAGTTCGTCAACGGT 389
DB 186 GTGGTTCGTCAGGATCGTCACCCCTGACACAGATGTTGCTGTGTACACCGGTG 127
QY 390 GCGCTCTTCGCGTCGCGCGTACGCCACACCGCGTTCGCCCTCGGCTTGACCGAGTT 449
DB 126 GATGTTCCACGCGCGGAGTTCCACAGCGTCACGTTTCGATATACACCGGTGTCGAGTG 67
QY 450 CAGCAGGCGGAGCAGTCGTCGCGGATGCCCTCGACGTCG 489
DB 66 CATCACTCGATCAGTACCCACGCGGTGACTTGAGCTTG 27

RESULT 39
US-10-259-165-267/c
; Sequence 267, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165

```

```

; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 267
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Oryza sativa
; NAME/KEY: misc feature
; LOCATION: (310)..(310)
; OTHER INFORMATION: k = guanine or thymine
; US-10-259-165-267

Query Match 10.6%; Score 52.8; DB 15; Length 1560;
Best Local Similarity 49.3%; Pred. No. 0.00062;
Matches 138; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 210 GATCTCGTCGGCGTGGTTCCAGGCGAAGAGCAGCGCGTAGTCCACCGGTGCGGCGTGAA 269
DB 847 GATCCGCTACTCGTCCACCGCGACTTGATCGCACCGCAGTCGTGCGCGGACACAGATGTA 788
QY 270 CGCGTCGGGTGCGCACCGGATGTGCTGCGCGGGGTGAGCCGCCCTGCTTGCGCGG 329
DB 787 GCAGTCTCTCGATCGCGAGCGTGGGAGCACAGTTCGATCGCACCGCAGTCGTGCGGCGCA 728
QY 330 CGTCGTGTCGCACACCCAGGAGACAGGTCCGGACCGATGCGCGCAGAGTTCGTCAACGGT 389
DB 727 GTGCGTCGTCGCGAGATCGTCACCCCTGACACAGATGTTGCTGCTGTACACCGGTG 568
QY 390 GCGCTCTTCGCGTCGCGCGCTAGCCACACCGCGTTGCCCTCGGCTTGAGCGAGTT 449
DB 667 GATGTTCCACGCGCGGAGTTCCACAGCGTCACGTTTCGATATACACCGGTGTCGAGTG 608
QY 450 CAGCAGGCGGAGCAGTCGTCGCGGATGCCCTCGACGTCG 489
DB 607 CATCACTCGATCAGTACCCACGCGGTGACTTGAGCTTG 568

RESULT 40
US-10-156-761-566/c
; Sequence 566, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 566
; LENGTH: 1044
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
; US-10-156-761-566

```



Query Match 10.5%; Score 52.6; DB 15; Length 1044;  
Best Local Similarity 46.7%; Pred. No. 0.00075;  
Matches 202; Conservative 0; Mismatches 229; Indels 2; Gaps 1;

QY 68 GTAGAGGTGGCGGGAAGCGGTACGCGCTTCGACGCTCAGCGGTTCAGCGGCGGCGGACACCGG 127  
Db 645 GTCCGCGCCGAAGCGCCAGGCGCGCGCTTCGCGGCGCTTCGCGCGGCTCAGCGCGATGAT 586  
QY 128 CGGGTCAGCTC--GTACAGTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCCCG 185  
Db 585 CGCGCGCGCGCGCGCGCGCGAGCGGTACGTTGTGCGCGCGACCTCCGCGCCACGCC 526  
QY 186 GCGGAACCTCTGCTCTTCGATGATCTCTGCTGCGGCTGCTTCCAGGCGAAGACGAGCGC 245  
Db 525 CCAGACGCGACGAGCTGGCGGCGCGCGCGCGCGCGGTGAGACGCGCGCGCGGTAGG 466  
QY 246 GTAGTCCACCGGTCCGCGCGGTGAACCGCTCCGCGGTGCGCACCGCGATGTGCGTCCGCG 305  
Db 465 GTGGAGACCGCGTCCGCGGATGATCGCGCGCTGTCGAGGCGAGGTGTCGCGGATGG 406  
QY 306 GGTGACCGCGCGCTGTCGCGCGGTGTCGCGACACCGCGAGGACGAGTCCGAGCC 365  
Db 405 GATGAGGCTGTCCTCCGCGCGCGCGGTGACTGGCGCGCGCGCGCGTGTGATGCGATGCC 346  
QY 366 GATGCGCGCAGAGTTCGTCAAGTGGCGGTCTTTCGCGCGCGCGTACGCCACACCGC 425  
Db 345 GCGGTGCGCATCTGAGTGCAGGGTCCGCTTGGCGAGCGCGCGCGCTGCGCGCAGGT 286  
QY 426 CTTGCCCTCGCGCTTTCGAGCGAGTTCAGCGGCGGAGCAGGTTCGTCGGATGCCCTCGAC 485  
Db 285 CTTGCCCGCGCTTCAGCGGTGACCGAGTACCGAGTACCGCGCGCGCTTTCGAGTCCGCGCG 226  
QY 486 GTCGCGCGCGAAC 498  
Db 225 GAGGCTGTGATC 213

RESULT 41  
US-10-156-761-6641/c  
; Sequence 6641, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6641  
; LENGTH: 1614  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1614)  
US-10-156-761-6641

Query Match 10.5%; Score 52.4; DB 15; Length 1614;  
Best Local Similarity 46.9%; Pred. No. 0.00078;  
Matches 196; Conservative 0; Mismatches 221; Indels 1; Gaps 1;

QY 66 GGTAGACGCTGGCGCGGAGGCGGTACGCGCTTCGACGCGTCCGCGCGGCGGACACC 125

Db 1493 GCGTAGACGACCTCGCGCCAGATCTTCGCGGCGTC-CTGGTGACCGCTTCGCGAGCTC 1435  
QY 126 GCGCGGTACGCTCGTACGCTGTCAGCGGCGGAGAGTACAGGATCCACTGTCCGCGAGCCCG 185  
Db 1434 CGGGTTCGTTTCGCGAGCGCGGATGTGAGAGCGGAGCGAACCGTAGTTCGCTGGCCAG 1375  
QY 186 GCGGAACCTCTGCTCTTCGCGCGATGATCTCTGTCGCGGTGTCGCGAGGAGAGAGCGCGC 245  
Db 1374 GTGCTTCGCGCTGTCGCGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1315  
QY 246 GTAGTCCACCGCTCGCGCGGTGAACCGCTCCGCGGTGCGCACCGGAGTGTGCGTCCGCGG 305  
Db 1314 GTGGCGACCGCGCGCGGTGTCGCGACACCGCGAGCGGCGAGTCTTTCGCGAGCTCGTT 1255  
QY 306 GGTGACCGCGCGCTGCTTCGCGCGGTGTCGCGACACCGCGAGGAGACCGAGGTCCGAGCC 365  
Db 1254 GATCGCTTCGAGTCTTCACCGCGGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1195  
QY 366 GATGCGCGCAGAGTTCGTACGCGGTGGCGCTTTCGCGGTGCGCGCGCGTACCGCGCGCG 425  
Db 1194 CGTACGCGCATGTGCGCGGAGGTGTCCTTCGCGCGCGCGCGCGCGCGCGCGCGCG 1135  
QY 426 CTTGCCCTCGCGCTTCGAGCGAGTTCAGCGGCGGAGCAGTTCGTCGGATGCCCTCG 483  
Db 1134 GCGCGCTTCGTCGACGCGCTTCGCGCTTTCGCGCTTCGAGGTGTCGCGCGCGCGCGCG 1077

RESULT 42  
US-10-156-761-6347/c  
; Sequence 6347, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 6347  
; LENGTH: 1182  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1182)  
US-10-156-761-6347

Query Match 10.4%; Score 52.2; DB 15; Length 1182;  
Best Local Similarity 48.3%; Pred. No. 0.00091;  
Matches 216; Conservative 0; Mismatches 218; Indels 13; Gaps 2;

QY 53 CCGCGGTAGTTCGCGGTGAGCGGTGGCGGCGGAGCGGTACGCGCTTCGAGCGTACGCGG 112  
Db 810 CACGAGGACGTACGCGCTCGTTCGCGCGCACAGGGTGGCGCGCGCGGTGATCCG 751  
QY 113 GCGGCGGACACCGCGCGGTTCAGTTCGTCAGCGTTCAGCGGCGGAGCGTACAGGATCCACT 172  
Db 750 GCGCGCGACTTCG-----TCGTACCGCGCACACAGGAGGTGTCGCGACTCGA 701  
QY 173 GTCGCGGACGCGCGGAGACTTCCTGCTTCGCGCATGATCTCTCGCGCGGTTCGAGG 232  
Db 700 TGTGTTGGCGCGCGAGCGTTCGATGTCGCTTCGCGTTCGCGCGCGGTGAGC 641

Qy 233 CGAAGACGACGGCGTAGTCACCGCGTCGGCGTGAAACGGCTCCGGGGTGCACCGGA 292  
Db 640 CGTCGAGCGAGGCGCGTGGCGACGCGCTGCGCGAGCGGATGGTGAGGACCGCGTCGA 581  
Qy 293 TGTGCGTCCGGGGGTAGCGCGGCGCTGCTTGGCGGCGGTCGTGTGCGCACACCGCAGGAGA 352  
Db 580 TCAGTGCAGCGCGTTTCAGCGCGC--GCTTGGCCCGCCAGGCGCGACGCGGTAGTACG 524  
Qy 353 CCAGGTCCGAGCGATGCGCGAGAGTTGTCAGCGTGGCGCTCTTCGCGGTGCGCGCGT 412  
Db 523 CCTCGTCCGCGGCGCCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 464  
Qy 413 ACGCACACCGCGTTCGCCCTCGCGCTTGAGCGAGTTTCAGAGGCGGCGAGCGTCCGTGC 472  
Db 463 CGGGGATCAGGTTCCGAGTCCCGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCT 404  
Qy 473 GGATGCCCTCGACGTCGCGGCGGAACC 499  
Db 403 GGCAGACGACTCCCGCGCGCGCGACCC 377

RESULT 43  
US-10-282-122A-25426  
; Sequence 25426, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR FILING DATE: 60/191,078  
; PRIOR APPLICATION NUMBER: 2000-03-21  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 25426  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Mycobacterium avium  
US-10-282-122A-25426

Query Match 10.4%; Score 52.2; DB 13; Length 1200;

Best Local Similarity 52.0%; Pred. No. 0.00091;  
Matches 117; Conservative 0; Mismatches 108; Indels 0; Gaps 0;  
Qy 266 TGAACGCGTCGGGGTGCACACCGGATGTGCTGCGCGGGGTGAGCGGCGCTGCTTGG 325  
Db 422 TGCTCAAGCGCGGGGTGTTCCACGACATCTCCGGGCGGTGATGCTGATCCCGGCGCG 481  
Qy 326 CGGGGTGCTGTCGACACACCGGAGACCGAGTCCGACCGATGCCGAGAGTTCTGTC 385  
Db 482 CGACATCGCGCGCGCGCTGCTGCGCTGTTCGAGGCGATGCTGGGCTACCGCGGCA 541  
Qy 386 CGGTGGCGCTTTCGCGCTGCGCGGTACGCGCAACCGCTTGCCCTGCGGCTTGAGCG 445  
Db 542 AGGAATCGACGCGCGCGCTGCGCGCGACCTTGGGCGTCAACGCGCTCGACGCGCTG 601  
Qy 446 AGTTTCAGCAGCGCGGAGAGGTGCGTGGGATGCCCTGACGCTCG 490  
Db 602 TCGCGAAGTGCCTGCGGCTGCTGCTCAGCAACTGGCGCGG 646

RESULT 44  
US-10-156-761-5628/c  
; Sequence 5628, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 5628  
; LENGTH: 1425  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1425)  
US-10-156-761-5628

Query Match 10.4%; Score 52; DB 15; Length 1425;  
Best Local Similarity 48.6%; Pred. No. 0.00099;  
Matches 142; Conservative 0; Mismatches 150; Indels 0; Gaps 0;  
Qy 194 CTGTCTCTTCGCCATGATCTGTCGGCGTGGTTCAGCGGAGAGCGCGGTAGTCA 253  
Db 1363 CCAGGAGCGTCTCACCAGCGGTCCCGCATGCCCGCGGCGACCGGGCCAGGA 1304  
Qy 254 CGCGTCCGGCGCTGAACCGGTCCGGGTGCGCACCGGATGTGCGTCCGGGTGAGCC 313  
Db 1303 GCGAGCGGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1244  
Qy 314 GCGCCTGTGTCGGCGCGTGTGTCGCAACCCAGGAGACCGAGTCCGAGCGCGCG 373  
Db 1243 CG 1184  
Qy 374 AGAGTTTCGTCAGGTGGCGCTTTCGCGGTGCGCGCGCGCGCGCGCGCGCGCG 433  
Db 1183 GCGAGTGGAGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1124  
Qy 434 CGGCTTTCAGCGAGTTTCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485  
Db 1123 CCAGGTCCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1072

```

; SOFTWARE: facsimile version 3.10
; SEQ ID NO 14
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-14

Query Match          10.2%; Score 51.2; DB 13; Length 975;
Best Local Similarity 48.0%; Pred. No. 0.0017;
Matches 146; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY      190  AACTCTGTCTCCCTTCGCCATGATCTCGTCGGCGTGTTCCAGGCGAAGACGACGGCGGTAG 249
Db      674  AGACACGCTTCCGACACGCTCGGCACGTCGACGTAGTCCCGGTGGCGCGACGTGGAC 615

QY      250  TCCACGCGGTCTCGGGCGGTGAACGCGTCTCGGGGTTCGGCACCGGGATGTGCTGTCGCGGGGTG 309
Db      614  AGCTTCCACACACCGCGGGAGGGTGTCCGGGGCCACGGCTCTCGCGAGCCGCCCGGCCACCC 555

```

QY 310 AGCGGCGCTGCTGGCGGGTGTGTCGCACACACCCAGGACCGAGTCCCGACCGATG 369  
Db 554 CGGCCCAGCAGCTGACCGCGGGGTGCCGGGCCCGGACATGCGGATCCGACGACG 495  
QY 370 CCGCAGAAGTTCGTACAGCGTGGCGCTCTTCGCGGTGCGCGGACACCCCGCTTG 429  
Db 494 CCGCGCTCGACGACGCGCTCGCGGGCGGCTCCAGCAGCGCGGCTCGCGCGAGCTTG 435  
QY 430 CCCTCGGCTTCAGCGAGTTCAGCAGGCGGAGCGAGTCCGCGGATGCCCTCGAGCTCG 489  
Db 434 GCCTTCGCGTAGTGGTTCGCGCGGGCGGCGTCCGCGTCCGCGGTCTCGCGGGC 375  
QY 490 GCGG 493  
Db 374 GCGG 371

## RESULT 48

US-10-282-122A-11252  
; Sequence 11252, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; PRIORITY FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 11252  
; LENGTH: 2010  
; TYPE: DNA  
; ORGANISM: Burkholderia cepacia  
US-10-282-122A-11252

Query Match 10.2%; Score 51.2; DB 13; Length 2010;  
Best Local Similarity 53.3%; Pred. No. 0.0015;  
Matches 130; Conservative 0; Mismatches 113; Indels 1; Gaps 1;  
QY 24 AACGAGTCGAGATACGCGGAGGAGAACACCCCGCGTAGTCCGCGGTAGACGCTGGCGCG 83

Db 1612 AACCTGCATACCGACGGCTATACGGAACCGCGCGCGCGCTGCGCGCGCGGC 1671  
QY 84 AAGCGGTACGGCGCTTCGACCGTACGCGGGGGGGGAGACACCGCGGGTCACTCTGTAC 143  
Db 1672 GAGACG-ACCCACGCTCGGCTTCTCGACGCTCGGCTCGGCTCGGCGGTGCGAGCTCGGCTC 1730  
QY 144 GTGTACGCGGGGGGAGTACAGGATCCACTGTCCGCGACGCGCGGGGAACCTCTGTCTCTT 203  
Db 1731 GATCGCAGCGCGCAGTTTACCGCGCGGGGACGCTCGGCTGGCGACATGCTTCCGCA 1790  
QY 204 GCCCATGATCTCTCGGCGGTGTTCCAGGCGGAGAGACGCGGTAGTCCACCGCGTGGG 263  
Db 1791 CGTGGCGCGCTCGTGGCGGTTTCACTGTCGGAATGGCGGCGACGCTGTTCCAGTGTGGG 1850  
QY 264 CGTG 267  
Db 1851 CGTG 1854

## RESULT 49

US-09-922-683-3/c  
; Sequence 3, Application US/09922683  
; Publication No. US20020192793A1  
; GENERAL INFORMATION:  
; APPLICANT: DECKER, Heinrich  
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
; PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS  
; GLA.O AND THEIR USE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/922,683  
; FILING DATE: 07-Aug-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/194,905  
; FILING DATE: 1999-12-01  
; APPLICATION NUMBER: DE 19622783.6  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granados, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 026083/0193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-922-683-3

Query Match 10.1%; Score 50.6; DB 9; Length 546;  
Best Local Similarity 49.8%; Pred. No. 0.0026;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;  
QY 239 GCAGCGCGTAGTCCACCGCGTCCGCGGTGAACGCGTCCGCGGTGCGCACCGGATGTGCG 298

Db 383 CGAGGAGCGGTACACCTCGTGGAGACGTGACGAGGGTGGCCACACCGTGGCGGA 324  
Qy 299 TCCCGGGGTGAGCGCGCCCTCTTGGCGGGTCTGTGTCGCACACCCAGGAGACCGGT 358  
Db 323 GCGCGCGTCGAGCAGGACCTGGTGGCCAGCAGTGTGGTGGGTGAAGGACCGCTGT 264  
Qy 359 CGCGACCGATGCCGAGAAAGTTGTCACGGTGGCGTCTTCGCCGTGCGCGGTACGCCA 418  
Db 263 CGGTGATGAGCGGTGACGAGTGGCACTCGGCCGGAAGTGCACGATGTCTGTGCGCGG 204  
Qy 419 CCACCGCTTGCCTCGGCTTGAGCGAGTTACAGAGGGGAGCAGGTGCGTGGATGC 478  
Db 203 CGGCCAGCGTGTGACGAGCGGGTGTGACACGTCGCCCTGGACGAAGGTGAGGCGG 144  
Qy 479 CCTCGACGTGCGCGCG 495  
Db 143 GATGTCACGACCGCG 127

## RESULT 50

US-10-156-761-2762/c  
; Sequence 2762, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2762  
; LENGTH: 1260  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1260)  
US-10-156-761-2762

Query Match 10.1%; Score 50.6; DB 15; Length 1260;  
Best Local Similarity 47.9%; Pred. No. 0.0022;  
Matches 146; Conservative 0; Mismatches 159; Indels 0; Gaps 0;  
Qy 188 GGAATCTCTCTCTGTCGCCATGATCTGTCGGCGTGTTCAGGCGAAGAGCGCGT 247  
Db 592 GGTAACGCGCGACCTCGTCCGCGATCGGGGGAACCGCGGAAGTCCAACTGGCGCGGT 533  
Qy 248 AGTCCACCGCGTGGGGGTGAACGCGTCCGGGTGGCGACCGGGATGTGTCGCCGGGG 307  
Db 532 ACGCGACAGCGCGGCCACGATCAGCTTCGGCTGGACCTCTTGGCCAGCTTCTGACCT 473  
Qy 308 TGAGCGCGCCCTGCTTGGCGGGTGTGTGTCGACACCCAGGAGACAGGTCCGGACCGA 367  
Db 472 CGGCATGTCCACCTGGCGGGTGTGTCGTCACGTGTGTCACGCGACGACGTTGTACAGT 413  
Qy 368 TGCGCAGAAAGTTGTCACGCGTGTGTCGCCGTCTTCGCCGTGCGCGGTACGCCACCCCGGT 427  
Db 412 TGCGGAGAAAGTTGATCTTCATGCCGTGGGTGAGTGCCTGCGCGGTTCAGAC 353  
Qy 428 TGCCCTCGGCTTGAGCGAGTTACAGAGGCGGAGCGAGGTGCGGTGCGGATGCCCTCGACGT 487  
Db 352 CCATGATCGTGTGTCGCCGGGTGAGCAGCGGAACATGGCGCGCGGTGGCTGCGCGG 293

Qy 488 CGGCG 492  
Db 292 CCGAG 288

Search completed: June 27, 2004, 21:02:34  
Job time : 360.804 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:12 ; Search time 2028.43 Seconds  
(without alignments)  
7360.905 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_1\_500

Perfect score: 500

Sequence: 1 ggtaccgaccgtgtccgg.....tcgagtcggcggaacct 500

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estlin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_pbg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	63.2	12.6	925	CNS0091P	AL053013 Drosophil
c 2	61.8	12.4	982	BX415111	BX415111 BX415111
c 3	61.6	12.3	982	BX415111	BX415111 BX415111
c 4	61.4	12.3	623	BQ295055	BQ295055 WHE2857_G

RESULT 1

CNS0091P/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

Drosophila melanogaster genome survey sequence TET3 end of BAC #

fly), genomic survey sequence.

AL053013

AL053013.1

GSS.

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

Genoscope.

Submitted (02-JUN-1999)

BP 191 91006 Evry cedex - FRANCE (E-mail : secref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

ALIGNMENTS

925 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
fly), genomic survey sequence.  
AL053013  
AL053013.1  
GI:4934461  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
Genoscope.  
Submitted (02-JUN-1999)  
BP 191 91006 Evry cedex - FRANCE (E-mail : secref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).

AL066051 Drosophil  
AL053013 Drosophil  
AV623523 AV623523  
AV625284 AV625284  
AV619788 AV619788  
BM002630 1031105G0  
BF860172 963015G07  
AL228850 Tetradon  
BF860172 963015G07  
AL066742 Drosophil  
AL066742 Drosophil  
BF860173 963015G07  
BF860173 963015G07  
AG043040 Pan trogl  
CK027202 FCAS01882  
CA622308 wln.pk00  
AL066742 Drosophil  
AL066742 Drosophil  
AL066051 Drosophil  
BZ365371 ie15a07.9  
BZ365371 ie15a07.9  
CC34644 OQAM79TH  
CG275728 OXFI14TH  
AG030607 Pan trogl  
BQ646304 AGENCOURT  
AL820302 AL820302  
CG324766 OQWB31TV  
CC662995 OGUJ523TH  
BX384405 BX384405  
BX384405 BX384405  
CB678034 OSUNEB15J  
AG074680 Pan trogl  
BQ080807 1030007C0  
BF630057 HVSM5B000  
BX425797 BX425797  
BX425797 BX425797  
CB678003 OSUNEB15I  
CB651766 OSUNEB16N  
CB672263 OSUNEB06D  
CB671209 OSUNEB04K  
BX442207 BX442207  
AL538546 AL538546  
CA075606 SCULAM106  
CG262283 CG2CH86TH  
CC443238 PUHEM44TD  
CG268544 OXAZ444TV  
CA758758 BR0300040  
CG47961 OGA9A78TH  
CG338271 OGCG91TH  
AG171124 Pan trogl  
CG659824 OSJNB01B  
CG652012 PUJFJ27TD  
BZ359621 id79c11.b





http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0CAP004BE10QPI.  
Location/Qualifiers  
1. .982  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0CAP004Y120"  
/tissue\_type="THYMUS"  
/note="Vector: PCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the PCMVSPORT 6 vector.  
Library was not normalized."

## ORIGIN

Query Match 12.3%; Score 61.6; DB 13; Length 982;  
Best Local Similarity 27.1%; Pred. No. 6.1;  
Matches 92; Conservative 100; Mismatches 148; Indels 0; Gaps 0;  
QY 81 GCGAAGCGTACGCGCTTCGAGGTACAGCGCGCGCGGACACCGCGGGTCACTCGT 140  
Db 451 SCGGCGGCCCGCCBCSCCGSCCGSCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 392  
QY 141 CAGGTGTACGCGGGGACGTACAGATCCACTGTCCGCCAGCGCGCGCGGAACTCCTGCTC 200  
Db 391 CCCCCSSSSCCCGSSSSCGSSSSCGSSSSCGSSSSCGSSSSCGSSSSCGSSSSCG 332  
QY 201 CTTGCGCATGATCTGTCGGGTGTTTCAGGGAGAGACGAGCGGTAGTCCACCGCTC 260  
Db 331 SCSSSSSSCCCGSSSSCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 272  
QY 261 GGGCGTGAACCGCTCGCGGGTGGCCACCGGGATGTCGTGCGGGGGTGAAGCGCGCTG 320  
Db 271 GGGGGCCCCGG 212  
QY 321 CTTGGCGCGCGTGTGTGACACCCAGAGACAGGTTCGACCGATCCCGAGAGTTT 380  
Db 211 CCCCCCGGGGGGGGCGGGGCGSSSSCGSSSSCGSSSSCGSSSSCGSSSSCGSSSS 152  
QY 381 CGTACGCTGGCGCTCTTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420  
Db 151 SSS 112

RESULT 4  
BQ295055/c  
LOCUS  
DEFINITION  
WHE2857\_G05\_N09ZS Wheat unstressed root tip cDNA library  
aestivum cDNA clone WHE2857\_G05\_N09, mRNA sequence.  
BQ295055  
BQ295055.1 GI:20810539  
EST.  
Triticum aestivum (bread wheat)

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Poideae; Triticeae; Triticum.

## REFERENCE

## AUTHORS

Anderson, O.D., Chao, S., Chin, A., Close, T.J., Crossman, C.,  
Gustafson, P., Lazo, G.R., Pham, J., Rausch, C.J., Ross, K., Wilson, C.  
and Woo, J.

## TITLE

The structure and function of the expressed portion of the wheat  
genomes - Unstressed root tip cDNA library

## JOURNAL

## COMMENT

Contact: Olin Anderson  
US Department of Agriculture, Agriculture Research Service, Pacific  
West Area, Western Regional Research Center  
800 Buchanan Street, Albany, CA 94710, USA  
Tel: 5105595773  
Fax: 5105595818  
Email: oanderson@pw.usda.gov

Sequences have been trimmed to remove vector sequence and low  
quality sequence with phred score less than 20  
Seq primer: SK primer.

## FEATURES

## source

1. .623  
Location/Qualifiers  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="Chinese Spring"  
/db\_xref="taxon:4565"  
/clone="WHE2857\_G05\_N09"  
/tissue\_type="Root tip at 1.0 to 1.5 mm stage"  
/dev\_stage="Four-day old seedling"  
/lab\_host="E. coli SOLR"  
/clone\_lib="Wheat unstressed root tip cDNA library"  
/note="Vector: lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown under  
hydroponic conditions for four days. Root tips were  
excised and snap frozen (Ross and Gustafson) and total RNA  
was prepared at University of Missouri, Columbia. Poly(A)  
RNA was purified, a cDNA library was made, and the cDNA  
clones were in vivo excised to give pBluescript  
SK(-) phagemids in the U3 Close lab (Chin and Close) at the  
University of California, Riverside. Plasmid DNA  
preparations and DNA sequencing were performed in the OD  
Anderson lab (all other authors)."

## ORIGIN

Query Match 12.3%; Score 61.4; DB 13; Length 623;  
Best Local Similarity 45.8%; Pred. No. 6.3;  
Matches 155; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
QY 189 GAATCTCTGCTTCCTTCCTCCATGATCTCGTGGCGTGTTCAGCGGAGACGCGGTA 248  
Db 357 GAGTCGTCTCTCCCGCAGACCCCTCCGGGTCATGTGGCCCTCGACCTCGCGGC 298  
QY 249 GTCCACCGCTCGCGCGCTGAACGGTCCGGGTGCGACCGGATGTGCTGCCGGGGT 308  
Db 297 CCCGACCGCTCGCGGGCGTGCAGTAGGCGAAGGGCGGCGGTAGGAGCGGAGCG 238  
QY 309 GAGCGGGCCCTGCTGTGGCGCGCTGTCGTCACACCCAGGAGACGAGTCCGACCGAT 368  
Db 237 CGCGGGATCGCGCGCGGCGCATGCGCGGACGACGAGAGCCATGCGCGGTCGCG 178  
QY 369 CCGCAGAGTTCGTCAGGTGGCGCTTCGCGCGTCCGCGCGCTACGCCACCGCGTT 428  
Db 177 GACGCGCGCGCGGCGGCGAGGCTGGTGAAGCTGCGCTCGTGGGCTCCACGCCCTG 118  
QY 429 GCCCTCGGCTTGAGCGAGTTTCAGCGCGGCGAGCAGGTGGTGGGATGCCCTCGAGTC 488  
Db 117 CGCTTCATGCGCGCGAGATCTCGAAGCGCGCGCGCGCGCGCGCTCGCGCGCG 58  
QY 489 GCGCGCGAACC 499  
Db 57 CGCGTGGAGC 47

## RESULT 5

## CNS006XK/c

## LOCUS

## DEFINITION

CNS006XK  
Drosophila melanogaster genome survey sequence T7 end of BAC #  
BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AL066051.1 GI:4945019  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

## AUTHORS

## TITLE

1 (bases 1 to 935)  
Genoscope.  
Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES  
source  
1..935  
    /organism="Drosophila melanogaster"  
    /mol\_type="Genomic DNA"  
    /db\_xref="taxon:7227"  
    /clone="BACR14N09"  
    /clone\_lib="RPCL-98"  
    /note="end : T7"

ORIGIN

Query Match                 12.3%; Score 61.4; DB 29; Length 935;  
Best Local Similarity   25.3%; Pred.No. 6.6;  
Matches 108; Conservative 122; Mismatches 197; Indels 0; Gaps 0;

QY   66 GGGTAGACGGTGGGCGGAAGCGCTACGCCTTCACGGTCAGGGGCGGCGGACACC 125  
     |||:::||:::||:::||:::||:::||:::||:::||:::||:::||:::||  
DB   935 GASCGGGGGSSSGSGSCGSGSGSGSCCCSGSCCSCGCCSCSCSSSSCCC 876  
  
QY   126 GCGCGGTTCAGCTCGTCACTGTCACGGGGGACGTACAGATCCAATGTCGCCAGCCG 185  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB   875 SSCGCGCGCGCGSGSSCCCGCGSSSGCGCCCGCGCGCGCGSCCGSCCSSC 816  
  
QY   186 GCGGAATCTTGCTCTTCGCGCATGATCTGTTGGGCGTGTTCAGGCGAAGACAGCGC 245  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB   815 CCG 756  
  
QY   246 GTATGTCCACCGCTCGGCGGTAAACGGCTCGGGGTGGCACGGGATGTGGTCCCGG 305  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB   755 GSGCSSSSGSGSGSGSGSGSGSGSCSCSGCSGSGSGSGSGSGSGSGSGSG 696  
  
QY   306 GGTGAGCCGCGCTCTTGGCGCGCGTGTGTGTCGAACCCAGAGAACAGGTTCGGACC 365  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB   695 CGGCG 636  
  
QY   366 GATCGCGGAGAATTGTCACGGTGGCGCTCTTCGCGTGGCGCTACGCCACACCG 425  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB   635 GCCCG 576  
  
QY   426 CTTCGCCCTCGGCTTAGCGAGTTTCAGCAGGGGAGCAGTCCGTCGCGATCCCTCGAC 485  
     ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  
DB   575 GSASGCGCGMGCRAGSGKMGASGSRCGGACSGSGGSBSRSRKKGASCACSCSAY 516  
  
QY   486 CTCGCGG 492  
     ::|::|::|  
DB   515 GKSGSGC 509

```
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BF 191 9106 EVRY cedex - FRANCE (E-mail : seqrf@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pierer de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
pi and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TEf3"

FEATURES             source
Query Match          12.2%; Score 61; DB 29; Length 925;
Best Local Similarity 13.9%; Pred.No. 7.5;
Matches 45; Conservative 157; Mismatches 122; Indels 0; Gaps 0;

QY    172   TGTCCGCGAGCGGGAACTCCTGCTTCCTCGCATGATCTGTCGGCGTGTTCCAG 231
Db     562   KGCSSGGBSCCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSCSC 621

QY    232   GCGAAGACGACGCGTAGTCCACCGCGTCGGCGGTGAACGCGTCGCGGGGTGCCACCGG 291
Db     622   SSKSVGCTSCSSSSSCSSSSSTSSSTSSSTSSSTSSSTSSSSSSSSSSSYTTKSTSASG 681

QY    292   ATGTGCGTCGCGGGGTGACCGCGCGTGTTCGCGCGCGTCTGTCGCACACCACGAG 351
Db     682   SWAGGGSGTGTSASSSSSTSSSSSVSSGSKSTBSGGBSSSGSSSSSTSSSBBS 741

QY    352   ACCAGTTCGGACCGATGCGCGAGAATTGCTCACGCTGGCGGTCTTCGCGCGTCCGCG 411
Db     742   CTSTSSSSSSSVSSSTCSCTCCSCSYSSSTSSSTSSSTSSSTSSSGSSSVGTSSSDST 801

QY    412   TAGCCACCAACCCTGTCCTCGGCTTCAGCGAGTTTCAGCAGGGCGACAGGTCCGTG 471
Db     802   STCCSCCCYMTCTCSYBMECYVTSCTGSSSSSGKGVTKCCTGCGCSSTNGMBSTSS 861

QY    472   CGGATGCCCTCGACGTGGCGGCG 495
Db     862   ACSSSSSSCSSSVSSSSKSSASS 895

RESULT 7
AV623523/c LOCUS AV623523 512 bp mRNA linear EST 15-DEC-2000
DEFINITION AV623523 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
reinhartii cDNA clone LC065b05_r 5', mRNA sequence.
ACCESSION AV623523
VERSION AV623523.1 GI:10772700
```

**KEYWORDS**  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

**REFERENCE**  
**AUTHORS** Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.  
**TITLE** Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii  
**JOURNAL** DNA Res. 7 (5), 305-307 (2000)  
**MEDLINE** 20539644  
**PUBMED** 11089912  
**COMMENT** Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

**FEATURES**  
 source  
 1..512  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

**ORIGIN**  
 Query Match 12.0%; Score 59.8; DB 9; Length 512;  
 Best Local Similarity 48.7%; Pred. No. 11;  
 Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGGTACGTCGTACGCGGGGAGCTACAGATCCACTGTCGCCAGCC 183  
 Db 373 CCCAGGAGGAGTCTATGAGGGGATGGCAGCGCTGCTGATGGACAGCGGCCAGC 314  
 QY 184 CGCGGAACTCTGCTTCTGGCCATGATCTGTCGGGCTGGTTCACGCGAAGAGCAGC 243  
 Db 313 CGCGCGCAGACACCTTCGGCAGCGGTGAGTCTCTTGGCGAAGTCGGGGTCCACGACG 254  
 QY 244 GCGTAGTCCACCGCTCGGGGCTGCGGCTCGGAGTGGCGACCGGGATGTCGTCGG 303  
 Db 253 GAGGCGAGCTCGGGGTGCGCTCGTAGGCGACCGCGATCGGCTCCAGGACTTCGCGCGG 194  
 QY 304 GGGGTGAGCGCGCCCTCTGTCGGCGCTGTCGCGGCTGCGACCGGGATGTCGTCGG 363  
 Db 193 ATGATGAGCCACCTTCAGATGCGCGCAGCGCCCGCAGGTTCGACGTTCCACTTCATC 134  
 QY 364 CGATCGCGCAGAGTTCGTCAGGTGGCGCTCTTCGCGCTCGCGCTACGCCACCACC 423  
 Db 133 TCACGCTCTTGGCCTTGATGTTTCATGCTGCGCTGCGGTACGACGATCTTGTGGGG 74  
 QY 424 CGCTTGCCTCGGCTTGAGCGAGTTCAGCAGGCG 458  
 Db 73 TACAGCGCGCGGCGACGCTGTTGATCAGCTGGCG 39

**RESULT 8**  
 AV625284/c  
**LOCUS** AV625284 512 bp mRNA linear EST 15-DEC-2000  
**DEFINITION** AV625284 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LC090d07\_r 5', mRNA sequence.  
**ACCESSION** AV625284  
**VERSION** AV625284.1 GI:10774461  
**KEYWORDS** EST.  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

Chlamydomonadaceae; Chlamydomonas.  
 1 (bases 1 to 512)  
 Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.  
 Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii  
 DNA Res. 7 (5), 305-307 (2000)  
 20539644  
 11089912  
 Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

**FEATURES**  
 source  
 1..512  
 Location/Qualifiers  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
 /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

**ORIGIN**  
 Query Match 12.0%; Score 59.8; DB 9; Length 512;  
 Best Local Similarity 48.7%; Pred. No. 11;  
 Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGGTACGTCGTACGCGGGGAGCTACAGATCCACTGTCGCCAGCC 183  
 Db 506 CCCAGGAGGAGTCTATGAGGGGATGGCAGCGCTGCTGATGGACAGCGGCCAGC 447  
 QY 184 CGCGGAACTCTGCTTCTGGCCATGATCTGTCGGGCTGGTTCACGCGAAGAGCAGC 243  
 Db 446 CGCGCGCAGACACCTTCGAGATGCGCGCAGCGCTGCTGATGGACAGCGGCCAGC 387  
 QY 244 GCGTAGTCCACCGCTCGGGGCTGCGGCTCGGAGTGGCGACCGGGATGTCGTCGG 303  
 Db 386 GAGGCGAGCTCGGGGTGCGCTCGTAGGCGACCGGATGCGGTCCAGGAACTGCGCGCG 327  
 QY 304 GGGGTGAGCGCGCCCTCTGTCGGCGCTGTCGCGGCTGCGACCGGGATGTCGTCGG 363  
 Db 326 ATGATGAGCCACCTTCAGATGCGCGCAGCGCCCGCAGGTTCGACGTTCCACTTCATC 267  
 QY 364 CCGATCGCGCAGAGTTCGTCAGGTGGCGCTCTTCGCGCTCGCGCTACGCCACCACC 423  
 Db 266 TCACGCTCTTGGCCTTGATGTTTCATGCTGCGCTGCGGTACGACGATCTTGTGGCG 207  
 QY 424 CGCTTGCCTCGGCTTGAGCGAGTTCAGCAGGCG 458  
 Db 206 TACAGCGCGCGGCGACGCTGTTGATCAGCTGGCG 172

**RESULT 9**  
 AV619788/c  
**LOCUS** AV619788 532 bp mRNA linear EST 07-NOV-2001  
**DEFINITION** AV619788 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LC015e02\_r 5', mRNA sequence.  
**ACCESSION** AV619788  
**VERSION** AV619788.1 GI:10768963  
**KEYWORDS** EST.  
**SOURCE** Chlamydomonas reinhardtii  
**ORGANISM** Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

**REFERENCE**  
**AUTHORS** Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y., and Tabata, S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii

JOURNAL DNA Res. 7 (5), 305-307 (2000)

MEDLINE 20539644

PubMed 11089912

COMMENT Contact: Erika Asamizu  
The First Laboratory for Plant Gene Research  
Kazusa DNA Research Institute  
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES  
Location/Qualifiers  
1..532  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="C9"  
/db\_xref="taxon:3055"  
/clone="LC015e02\_r"  
/clone\_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"  
/note="Vector: pBluescript II SK-; Site: 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

ORIGIN

Query Match 12.0%; Score 59.8; DB 9; Length 532;  
Best Local Similarity 48.7%; Pred. No. 11;  
Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGTCAAGTCGTACCGGGGACGTACAGATCCACTGTCCGCCAGCC 183  
DB 375 CCCAGGAGGAGTCAATGAGGGATGGCAGCGCTCGTGATGGACGCGGCCACG 316

QY 184 CGCGGAACCTCTGCTCCATGATCTCGTCCGGTGTTCACGCGGAGAGCAGC 243  
DB 315 CGCGCCAGGACGCTCGGACGGTGTGAGTCTTGGCGAAGTCGGGGTCCACAGCAGC 256

QY 244 CGGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGACCGGATGTCGTCCCG 303  
DB 255 GAGGACACTCGGGTTCGGTCTGTAGCGACGCGATGCGGTCCAGMACTGCGCGCG 196

QY 304 GGGGTGACCGGCGCTTGGCGCGGTGTGTGCGACACCCAGGACAGAGTCCGGA 363  
DB 195 ATGATGACGACCACTTCAGATCGCGCCAGCGCGCCAGTGCAGTTCACATTCATC 136

QY 364 CGGTGCGCGAGAGTTCGTACGCTGGCGCTCTTCGCCGCTCGCGCTACGCCACCA 423  
DB 135 TCCAGCTCTTGGCTCTGATATGTTATGTCCTGCGGTACGAGCAGATCTTGCTGG 76

QY 424 CGCTTGCCTCGGCTTGAAGCGAGTTCAGCAGGCG 458  
DB 75 TACAGCGCGCGGCGCACGCTCGTTGATCAGCTGCGC 41

RESULT 10  
BM002630/c  
LOCUS 599 bp mRNA linear EST 25-OCT-2001  
DEFINITION 1031105G06.y1 C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BM002630  
VERSION BM002630.1 GI:16437410  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.  
AUTHORS Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants. Project: 1031  
JOURNAL Unpublished (2001)

COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

FEATURES  
Location/Qualifiers  
1..599  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Stress II (normalized), Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; Stress condition II library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP (NH4+ - containing) and shifted to TAP - NO3- (24hrs); H2 production conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP + sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al., (1996) Genome Research 6: 791-806."

ORIGIN

Query Match 12.0%; Score 59.8; DB 12; Length 599;  
Best Local Similarity 48.7%; Pred. No. 11;  
Matches 163; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 124 CCGCGGGTCAAGTCGTACCGGGGACGTACAGATCCACTGTCCGCCAGCC 183  
DB 495 CCCAGGAGGAGTCAATGAGGGATGGCAGCGCTCGTGATGGACGCGGCCACG 436

QY 184 CGCGGAACCTCTGCTCCATGATCTCGTCCGGTGTTCACGCGGAGAGCAGC 243  
DB 435 CGCGCCAGGACGCTCGGACGGTGTGAGTCTTGGCGAAGTCGGGGTCCACAGCAGC 376

QY 244 GGTAGTCCACCGCTCGGGGTGAACCGCTCCGGGTGCGACCGGATGTCGTCCCG 303  
DB 375 GAGGACACTCGGGTTCGGTCTGTAGCGACGCGATGCGGTCCAGMACTGCGCGCG 316

QY 304 GGGGTGACCGGCGCTTGGCGCGGTGTGTGCGACACCCAGGACAGAGTCCGGA 363  
DB 315 ATGATGACGACCACTTCAGATCGCGCCAGCGCGCCAGTGCAGTTCACATTCATC 256

QY 364 CGGTGCGCGAGAGTTCGTACGCTGGCGCTCTTCGCCGCTCGCGCTACGCCACCA 423  
DB 255 TCCAGCTCTTGGCTCTGATATGTTATGTCCTGCGGTACGAGCAGATCTTGCTGG 196

QY 424 CGCTTGCCTCGGCTTGAAGCGAGTTCAGCAGGCG 458  
DB 195 TACAGCGCGCGGCGCACGCTCGTTGATCAGCTGCGC 161

RESULT 11  
BF860172/c  
LOCUS 786 bp mRNA linear EST 19-JAN-2001  
DEFINITION BF86015G07.y1 C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF860172  
VERSION BF860172.1 GI:12250309  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii

Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.

## REFERENCE

1 (bases 1 to 786)

## AUTHORS

Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,

## TITLE

Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D.

## JOURNAL

Analyses of the Chlamydomonas reinhardtii Genome: A Model.

## COMMENT

Unicellular System for Analyzing Gene Function and Regulation in

## JOURNAL

Unpublished (2000)

## COMMENT

Contact: Charles Hauser

## JOURNAL

DCMB Box 91000

## COMMENT

Durham, NC 27708-1000

## JOURNAL

Tel: 919 613 8159

## COMMENT

Fax: 919 613 8177

## JOURNAL

Email: chause@duke.edu.

## COMMENT

Location/Qualifiers

## JOURNAL

1. .786

## COMMENT

/organism="Chlamydomonas reinhardtii"

## JOURNAL

/mol\_type="mRNA"

## COMMENT

/strain="CC-1690 wild type mt+ 21gr"

## JOURNAL

/db\_xref="taxon:3055"

## COMMENT

/clone\_lib="C. reinhardtii CC-1690, Stress condition I,

## JOURNAL

normalized, lambda Zap II"

## COMMENT

/notes="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2:

## JOURNAL

XhoI; This library, constructed by John Davies and Jeffrey

## COMMENT

McDermott, combines cDNAs from CC-1690 cells grown to

## JOURNAL

mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,

## COMMENT

1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min,

## JOURNAL

1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was

## COMMENT

purified from each sample, pooled and cDNA synthesized.

## JOURNAL

The cDNA was directionally cloned into lambda Zap II

## COMMENT

(Stratagene) in the EcoRI (5') and XhoI (3') sites.

## JOURNAL

pBluescript II SK- plasmids were excised from the lambda

## COMMENT

ZAP clones by superinfection with ExAssist (Stratagene)

## JOURNAL

phase. The library was normalized using method 4 described

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

## JOURNAL

in Bonaldo et al (1996) Genome Research 6: 791-806."

## COMMENT

in Bonaldo et al (1996) Genome Research 6: 791-806."

QY 492 GCGAACCT 500  
 Db 350 GGTGTCAT 358

RESULT 13  
 CNS0072Q/c  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #  
 BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL066742  
 VERSION AL066742.1 GI:4945205  
 KEYWORDS GSS.  
 ORGANISM Drosophila melanogaster (fruit fly)  
 SOURCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 932)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aarton Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 Location/Qualifiers  
 1..932  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR14B09"  
 /clone\_lib="RPCI-98"  
 /note="end : T7"

ORIGIN  
 Query Match 11.8%; Score 59; DB 29; Length 932;  
 Best Local Similarity 29.3%; Pred. No. 15;  
 Matches 139; Conservative 104; Mismatches 232; Indels 0; Gaps 0;

QY 1 GGTACCCGACGCTGTCCTCCGGAACACGAGTCGAGATACGCGGAGGAGAACACCCCGGT 60  
 Db 907 GSGC-SGSGSGCCCGCCSSSCGCGCGGCGSGCGSGCGCGCGCGSGSGSGCGCGC 848

QY 61 AGTCGCGGTAGACGGTGGCGCGAGGCGTACGCGCTTCGACGCTACGCGGCGGCGG 120  
 Db 847 CGCSGGGCGCGSGGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGSG 788

QY 121 ACACGGGGGTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCTACGCT 180  
 Db 787 CGCGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSG 728

QY 181 GCCCGCGGGAACCTCTGCTCTCCATGATCTCGTCGGGTGTTTCAGCGGAAGAGC 240  
 Db 727 GSSGCGSGSGCGSGCGSGCGSGCGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 668

QY 241 ACGGCGTAGTCCACCGCTCGCGGCGTGAACGGTCCGGGTGCGGACCGGATGTGGTG 300  
 Db 667 CSSGCGCGSGGCG 608

QY 301 CGCGGGGTAGCGCGCCCTGCTTGGCGCGCGCTGCTGCGCACCCAGGACGAGTCC 360  
 Db 607 GSGCSCCGCGCMVARMVSVSCCCSCCMASCCCGCVSGCGSCCSKSCCCCGCA 548

QY 361 GGACCGATGCCCGACGAAGTTCGTACGGTGGCGCTCTTCGCGCTGCGCGGTACGCCACC 420  
 Db 547 SCVACGCMGACGCMGCGCMGCCGCCSCGSCCGCGCGCGCGCGCGCGCGCGCMVCGCM 488

QY 421 ACCCGCTTCCCTCGCGCTTACGCGAGTTCACAGCGGCGAGCAGGTTCGCTGGGA 475  
 Db 487 CCMGVACCMCGCGVGMGCMGRCVSSAGSGCAGMGARARACAGSGCAGM 433

RESULT 14  
 BF860173/c  
 LOCUS  
 DEFINITION 963015G07.y2 C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BF860173  
 VERSION BF860173.1 GI:12250310  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 750)  
 AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 3  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Charles Hauser  
 DCMB Box 91000  
 Duke University  
 Durham, NC 27708-1000  
 Tel: 919 613 8159  
 Fax: 919 613 8177  
 Email: chauser@duke.edu.

FEATURES  
 Location/Qualifiers  
 1..750  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="CC-1690 wild type mt+ 21gr"  
 /db\_xref="taxon:3055"  
 /clone\_lib="C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II"  
 /note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN  
 Query Match 11.8%; Score 58.8; DB 10; Length 750;  
 Best Local Similarity 48.8%; Pred. No. 16;  
 Matches 159; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 133 CAGCTCGTCACGCTACGCGGCGAGTACAGATCCACTGTCGCGACCGCGGGAAC 192  
 Db 567 CAGGTATGAGGGGCTGGGCGACACGCTCGTGTGATGACAGCGCGCGCGCGCG 508  
 QY 193 TCCTGCTCTTCGCCATGATCTGTCGGCGTGGTTCAGCGGAGAGAGCGGTAGTCC 252

```

Db      507  GAGCGCTCGGACAGCGGTGAGCTCTTTGGGAACTCGGGGTCCACAGCAGGAGGCGAGC 448
Qy      253  ACCGCTTCGGGCTGTAACGGTCCGGGTGCGCACCGGATGTGCTGCGGGGTGAGC 312
Db      447  TCGGGGTTCGGCTCGGTAGGCGAGCGGATGCGGTCCAGGAACCTGCGCGGGAATGATGCGAG 388
Qy      313  CGGCGCTTCGTCGGCGCGGTGCTGTGTCGCACACCCAGGAGACAGGTCCTCCGACCGATGCGG 372
Db      387  CCACCTTCAGATGCGGCGCAGCGCCGAGTCCAGCTTCACCTTCATCTCCAGCTC 328
Qy      373  CAGAAGTTCGTCAGGTGCGGCTCTTCGCGGTGCGGCTGAGCCACACCGCTTGCCC 432
Db      327  TTGGCGCTTGATAATGTTCAATGCTCCTCGCGGTACGAGCAGATCTTGTCGCTACAGCGCC 268
Qy      433  TCGGCGCTTCGAGCGAGTTTCAGCAGGCG 458
Db      267  GCGGCGACGTCGTTGATCAGTCGC 242

```

```

RESULT 15
AG043040/c
LOCUS   Pan troglodytes DNA, clone: PTB-021D18.R, genomic survey sequence.
DEFINITION
AG043040
ACCESSION
VERSION  AG043040.1 GI:16571765
KEYWORDS
SOURCE   Pan troglodytes (chimpanzee)
ORGANISM
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

```

```

REFERENCE
AUTHORS  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
          Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE    BAC end sequences of Library PTB
JOURNAL  Unpublished
AUTHORS  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
          Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE    Direct Submission
JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
          and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
          1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          (E-mail: chimpbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
          Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT  Clones are derived from the chimpanzee BAC library PTB. This BAC end
          was generated during the R&D process and may have higher chance of
          clone-tracking errors.
PRIMERS
          Sequencing: M13Rev
LIBRARY
          Vector      : pKS145
          R.site 1    : SacI
          R.site 2    : SacI
          Location/Qualifiers
            1. 918
              /organism="Pan troglodytes"
              /mol_type="genomic DNA"
              /db_xref="taxon:9598"
              /clone="PTB-021D18.R"
              /sex="male"
              /cell_type="lymphoblast"
              /clone_lib="PTB Chimpanzee Male BAC Library"

```

```

FEATURES
source
Query Match      11.7%; Score 58.6; DB 29; Length 918;
Best Local Similarity 46.2%; Pred. No. 17;
Matches 181; Conservative 0; Mismatches 211; Indels 0; Gaps 0;
Qy      53  CCCCCTGATCGGGTAGCGTGGCGCGGAGGCGGTACGCGCTTCGACCGGTGAGCGG 112
Db      836  CCCCCTGATCGGGTAGCGTGGCGCGGAGGCGGTACGCGCTTCGACCGGTGAGCGG 777

```

```

ORIGIN
Query Match      11.7%; Score 58.6; DB 29; Length 918;
Best Local Similarity 46.2%; Pred. No. 17;
Matches 181; Conservative 0; Mismatches 211; Indels 0; Gaps 0;
Qy      53  CCCCCTGATCGGGTAGCGTGGCGCGGAGGCGGTACGCGCTTCGACCGGTGAGCGG 112
Db      836  CCCCCTGATCGGGTAGCGTGGCGCGGAGGCGGTACGCGCTTCGACCGGTGAGCGG 777

```

```

Qy      113  CGCGGCGGACACACCGCGGCTCAGCTCGTACGCGGGACGTACAGGATCCACT 172
Db      776  GCGGCGCGGCGCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Qy      173  GTCCGCGAGCCCGCGGAACTCTGCTCTTTCGCGCATGATCTCGTCCGCGTGGTTCAGG 232
Db      716  GGCCNCCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
Qy      233  CGAAGAGCAGCGCGTAGTTCACCGCGTGGGCGGTGAACGGTCCGCGGTGCGCACCGGGA 292
Db      656  CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
Qy      293  TGTGCGTCCGCGGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 352
Db      596  CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 537
Qy      353  CCAGGTCCGAGCCGATGCGCGCAGAGTTCGTTCAGGTGGCGCTCTTCGCGCGCGCT 412
Db      536  GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 477
Qy      413  ACGCCACACCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444
Db      476  CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 445

```

```

RESULT 16
CK207202/c
LOCUS   1101 bp mRNA linear EST 08-DEC-2003
DEFINITION FGAS018821 Triticum aestivum FGAS: Library 5 GATE 7 Triticum
          aestivum cDNA, mRNA sequence.
ACCESSION CK207202
VERSION    CK207202.1 GI:39569592
KEYWORDS  EST.
SOURCE    Triticum aestivum (bread wheat)
ORGANISM  Triticum aestivum
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
          Poideae; Triticeae; Triticum.
          1 (bases 1 to 1101)
          Allard, F., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
          Genswein, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
          Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,
          Penniket, C., Roach, J.L. and Sarhan, F.
          Functional Genomics of Abiotic Stress In Wheat and Canola Crops
          Unpublished (2003)
          Contact: Wm L Crosby
          Bioinformatics
          University of Saskatchewan, Department of Computer Science
          1C101 Engineering Building, 57 Campus Drive, Saskatoon,
          Saskatchewan, S7N 5A9, Canada
          Tel: 306 966 1769
          Fax: 306 966 2033
          Email: fgas_est@cs.usask.ca
          This sequence is the direct result of the Base calling software
          Phred (default parameters). It is the raw base calls. To aid in the
          identification of the high quality insert the software Lucy
          (default parameters) has been run on this sequence. Lucy identified
          the region [39,775].
          Plate: LSB010 row: A column: 10.
          Location/Qualifiers
            1. 1101
              /organism="Triticum aestivum"
              /mol_type="mRNA"
              /db_xref="taxon:4565"
              /clone_lib="Triticum aestivum FGAS: Library 5 GATE 7"
              /note="Vector: pCMV.SPORT6; Crown and developmental stages
              of spike formation in wheat cultivar Norstar 4 mRNA
              populations were combined before constructing the library.
              The first mRNA population is from 1cm crown sections after
              30 days of cold acclimation. The second is from 1cm crown
              sections after 11 days of deacclimation (before
              deacclimation plants were fully vernalized for 49 days).
              The third is from different developmental stages of spike

```



formation (5 to 50mm) that still have not emerged from the leaf (dissection required). The last is from different developmental stages of spike and seed formation after having emerged from the leaf (visible). First strand synthesis in this library was done in the presence of methylated dCTP thereby protecting from internal cleavage with NotI."

ORIGIN

Query Match 11.7%; Score 58.6; DB 14; Length 1101;  
Best Local Similarity 49.3%; Pred. No. 18;  
Matches 151; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 189 GAATCTCTCTCTTCCGCGGATGCTCGTGGCGTGTTCAGCGGAGAGCGCGGTA 248  
DB 678 GAAATCACCAATCTTCCACGAGTTCTTTGTTGGTCCGCGCGGGGTGAGGGTCTTGAT 619  
QY 249 GTCCACCGCTCGCGCGGTAACGCGTCCGCGGTGCGCACCGGGATGTGCGTCCGCGGGGT 308  
DB 618 GCGGACAGTCTGATGAGGNGCAGCGCGGTCTCTCGGTGACACCGGGTGTGGAC 559  
QY 309 GAGCGGCGCTTGTGGCGCGTGTGTCGTCACACCGAGAGACCGATCCGAT 368  
DB 558 GACGACCTCGACCTCGTGGCGGTGGCGTCCAGCGCCAGAGTAGGAGTCCAGCGGTT 499  
QY 369 GCGGACAGAGTTGCTACGCTGGCGTCTTCTGCGCGTGGCGGTACCGCCACCGCGTT 428  
DB 498 GCTGCTGTACATGCTCTGATGACAGCACCGCGGATGCGCGGACAGCTTGAG 439  
QY 429 GCGCTCGCGCTGAGCGATTCAGACGGGCGAGCAGTGTGCGGATGCGCGACGTC 488  
DB 438 GCGCTCGCGCTGCGGCGACGCTCTCGCGCGCTGAAGGTGAGGCTGTAGCGCGCGGG 379  
QY 489 GCGGCG 494  
DB 378 GACGCG 373

RESULT 17  
LOCUS CA622308/c  
DEFINITION wlin.pk0089.d10 wlin Triticum aestivum cdna clone wlin.pk0089.d10  
5' end, mRNA sequence.  
ACCESSION CA622308  
VERSION CA622308.1 GI:2520604  
KEYWORDS EST.  
SOURCE Triticum aestivum (bread wheat)  
ORGANISM Triticum aestivum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.  
REFERENCE 1 (bases 1 to 581)  
Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z., Miao, G., Caraher, N. and Hanafey, M.K.  
DuPont Wheat cDNA Sequence  
Unpublished (2002)  
Contact: Scott V. Tingey  
Crop Genetics  
E. I. DuPont de Nemours and Company  
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA  
Tel: 302-631-2602  
Fax: 302-631-2607  
Email: Scott.V.Tingey@USA.dupont.com  
Seq primer: M13.  
Location/Qualifiers  
1..581  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/db\_xref="taxon:4565"  
/clone="wlin.pk0089.d10"  
/tissue\_type="leaf"  
/clone\_lib="wlin"  
/note="Vector: pBluescript SK+; Site\_1: EcoRI; Site\_2:

XhoI; Wheat (Triticum aestivum L.) leaf 7 day old seedling, light grown (normalized)"

ORIGIN

Query Match 11.6%; Score 58.2; DB 14; Length 581;  
Best Local Similarity 47.6%; Pred. No. 19;  
Matches 171; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 125 CGCGCGGTACGCTGTCACGTGACGCGGGAGCTACAGGATCCACTGTCGCGAGCCC 184  
DB 360 CGCGCGGATTTGCTGAGGCTGGGCTCGCTGACGACCTGCGCGAGCTCTTGGCGGCTCT 301  
QY 185 GCGGAACTCTCTCTTCCGCGGATGCTCGTGGCGTGTTCAGCGGAGAGCAGCG 244  
DB 300 CTTGACACAGTACCCCTTGTGATGTACTTCTGAAACGCGGGAGAGAGCGGAACA 241  
QY 245 CTTAGTTCACCGCTCGCGCGTGAACGCTGCGGGTGGCACCGGGATGTCGCGCGG 304  
DB 240 ACGGACGCTGTCAACGAGAGCGGCTGTGGCGATGACCCCGGTAGAGCGAGCGGA 181  
QY 305 GGGTGAAGCGCGCTTGTGGCGGCTGTGCTGCGCACACCGAGAGACCGAGTCCGGAC 364  
DB 180 AGTCAACGCGCTCTCTGCTGTTACCGCGGTGGAATCTCTGATGTGATGATATTCG 121  
QY 365 CGATGCGCGAGAGTTCGTCACGCTGGCGCTTTCGCGCTCGCGCTACCGCCACCGCC 424  
DB 120 ACACCTTGTCTCTCTTGTACGCTTGGCGCGCTCGAATCTCGCGCGCTCAATCATG 61  
QY 425 GCTTCCCTCGCGCTTGTGACGAGTTCAGACGGCGGCGAGCTGCGGTGCGGATGCG 483  
DB 60 CACTGCGCGCGCTGTACGCGCGCGCGCGCGCGCTCAGGTCTCCCGAGGTGGCGCTTG 2

RESULT 18  
LOCUS CNS0072Q  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL066742.1 GI:4945205  
VERSION AL066742  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 932)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's Pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Location/Qualifiers  
1..932  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"

FEATURES	Location/Qualifiers
source	1. .935
organism	"Drosophila melanogaster"
mol_type	"genomic DNA"
db_xref	"taxon:7227"
clone	"BACR14N09"
clone_lib	"RPCI-98"
note	"end : 17"

ORIGIN	Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
115	GGCGGACACCGCGCGGTACGCTGCTACGTTACGCGGGGACGTACAGATGTCACGTGT	174	11.6%	58	29	1	
495	VSTGTGTCMVMVCGCGSCMCRITSSGSGTSCGTCMVMSSVSCSCSGTCCGCGCSCTSC	554	28.5%	21			
175	CGCCAGCCCGCGGAACTCTGCTTCCTTCGTCATGATCTGTCGGGTGTTCCAGCGC	233					
555	KWCSTYGGCKGCGCGCTSCSSSSCCSBBSYSTCTCTKCSGGSTGSGCTGCCGGG	614					
235	AGAGCAGCGCTATTCACCGCTCGGCGGTGAACCGTCCGGGTTCGACACCGGGATG	294					
615	GSQCG	674					
295	TCGCTGCCGGGTGAGCGCGCGCTGCTGCGCGCGCTGCTGCGACACCGAGAGACC	354					
675	GSQCG	734					
355	AGTCCCGGACCGATCCGCGAGAGTTCGTACAGGTTCAGCGGTCTTTCGCGTCCGCGGTAC	414					
735	GCSCCG	794					
415	GCACACCGCTTGCCTTCGCGCGCGCTGAGCGAGTTCAGCGGGCGAGC-AGTCCGTGCG	473					
795	CSGCG	854					
474	GATCCCTCCAGCTCG	495					
855	SGSSCG	876					

RESULT 20	BZ365371	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	JOURNAL	COMMENT
BZ365371	1e15e07.g2 WGS-Zmaysf (JM107 adapted methyl filtered)	588 bp	DNA linear GSS 19-NOV-2000	BZ365371	1	GI:25074825	Zea mays	Zea mays	1 (bases 1 to 588)	Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L., Zutavern,T., McCombie,W.R. and Martienssen,R.A.	Genomic shotgun sequences from Zea mays (methyl-filtered) Unpublished (2002)	Contact: W. Richard McCombie Lita Annenberg Hazen Genome Sequencing Center Cold Spring Harbor Laboratory PO Box 100, Cold Spring Harbor, NY 11724, USA Tel: 516 367 8884 Fax: 516 367 8874 Email: mcombie@cshl.org Plate: 1e15 row: e column: 07 Seq primer: -21M13UnivRev Class: shotgun High quality sequence stop: 588.

ORIGIN	Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
57	CGGTAGTCGGGTAGACGTCGCGCGGAGCGGTACGCGCTTCGACGCGTACGCGGCGG	116	11.6%	58	29	1	
542	CTGBGTCGCGGGSXSGSGSCGBGCGGGSTKGGSGGGGBSBKTYTKGKGCGG	601	32.6%	21			
117	GGGACACCGCGGGTCACTGCTGCTACGTTACGCGGGGAGCTACAGATCCACTGTC	176					
602	SSGSCCGSGSCRGGGGCG	661					
177	GCAGACCGCGCGGAACTCTGCTTCCTTCGCTGATGATCTGCTGCGGTGTTCCAGGGG-A	235					
662	GGCSGGSGSGSGCG	721					
236	AGAGCAGCGGTAGTCCACCGCTCGGGGTGAACCGTTCGGGTTCGCGACCGGATGT	295					
722	GGCSGGCG	781					
296	GCCTGCCGGGTGAGCGCGCGCTGCTGCGCGCGGTGCTGCGACACCGAGAGACCA	355					
782	CGCGCGCGSGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	841					
356	GTTCGCGACCGATGCGCGAGAGTTCGTACGCGTTCGCGCTTCGCGCGTCCGCGCGTAC	415					
842	CSGCG	901					
416	CCACACCGCTGCTGCTGCGCGCTGAGCG 445						
902	SGCGCGCGSGSGSGCG	931					

ORIGIN	Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
127	Conservative	84	Mismatches	178			

ORIGIN	Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
57	CGGTAGTCGGGTAGACGTCGCGCGGAGCGGTACGCGCTTCGACGCGTACGCGGCGG	116	11.6%	58	29	1	
542	CTGBGTCGCGGGSXSGSGSCGBGCGGGSTKGGSGGGGBSBKTYTKGKGCGG	601	32.6%	21			
117	GGGACACCGCGGGTCACTGCTGCTACGTTACGCGGGGAGCTACAGATCCACTGTC	176					
602	SSGSCCGSGSCRGGGGCG	661					
177	GCAGACCGCGCGGAACTCTGCTTCCTTCGCTGATGATCTGCTGCGGTGTTCCAGGGG-A	235					
662	GGCSGGSGSGSGCG	721					
236	AGAGCAGCGGTAGTCCACCGCTCGGGGTGAACCGTTCGGGTTCGCGACCGGATGT	295					
722	GGCSGGCG	781					
296	GCCTGCCGGGTGAGCGCGCGCTGCTGCGCGCGGTGCTGCGACACCGAGAGACCA	355					
782	CGCGCGCGSGSGSGSGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	841					
356	GTTCGCGACCGATGCGCGAGAGTTCGTACGCGTTCGCGCTTCGCGCGTCCGCGCGTAC	415					
842	CSGCG	901					
416	CCACACCGCTGCTGCTGCGCGCTGAGCG 445						

## FEATURES

Location/Qualifiers

1. .588  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="tel5e07"  
 /lab\_hosts="JM107 or DH5a"  
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;  
 The vector was digested with XbaI and one nucleotide was  
 added by fill in the recessive 3' end. The genomic DNA  
 was nebulized, end repaired, adaptor ligated and size  
 fractionated using sephadex. The resulting fragments were  
 between 0.8 and 3 kb and were cloned into the vector  
 (.x/y reads in M13mp19, .b/g reads in pUC19). The same  
 ligation was transformed in either JM107 or DH5a."

## ORIGIN

Query Match 11.4%; Score 57; DB 28; Length 588;  
 Best Local Similarity 50.5%; Pred. No. 29;  
 Matches 138; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 188 GGAATCTCTGCTCTCCATGATCTGTCGGGCTGTTCCAGGCGAAGACGACGCGT 247  
 Db 131 GGTCCACACCTCCCGGTCGGACCTCTCTCGCGTCTCTCCACCTTGATGATCGT 190

QY 248 AGTCACCGCTCGGCGTGAACGCTCCGGGTCGCGACCGGATGTGTCGCGGGG 307  
 Db 191 AGTAGCATGTTGGGACGAGCGGTAGGCGACGTCTCTCGCGCGTCCAGCGGGGGA 250

QY 308 TGAGCGGCGCTGCTTGGCGCGCTGCTGTGCGCACACCGAGGACGAGTCCGACCGA 367  
 Db 251 GTTGATGCCGAGTAGCACTCCGAGGAGGCGTACATGTTGACACGAGCGGCGCGC 310

QY 368 TGCGCAGAGTGTGTCAGGTGGCGCTCTTCGCGCTGCGCGCTACGCCACACCGCGT 427  
 Db 311 CGCGTAGACTCCAGCGGGCGACGTACTCGCGCATGACCCCGTACGATGACGTCGA 370

QY 428 TGCCCTCGGCTTGAGCGAGTTCAGCAGGGCGGA 460  
 Db 371 TGTACTTGTGCGCGCCAGAGCGCGCGACGA 403

## RESULT 21

CC334644  
 LOCUS  
 DEFINITION  
 OGA079TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0367M13,  
 genomic survey sequence.

CC334644  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

CC334644.1 GI:30804057  
 GSS.

## ORGANISM

Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

1 (bases 1 to 755)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

## TITLE

Consortium for Maize Genomics

## JOURNAL

Unpublished (2002)

## COMMENT

TIGR  
 Contact: Cathy Whitelaw

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .755

## ORIGIN

Query Match 11.4%; Score 57; DB 28; Length 755;  
 Best Local Similarity 50.5%; Pred. No. 29;  
 Matches 138; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 188 GGAATCTCTGCTCTCCATGATCTGTCGGGCTGTTCCAGGCGAAGACGACGCGT 247  
 Db 336 GGTCCACACCTCCCGGTCGGACCTCTCTCGCGTCTCTCCACCTTGATGAACTCGT 395

QY 248 AGTCACCGCTCGGCGTGAACGCTCCGGGTCGCGACCGGATGTGCGCTGCCGCGGG 307  
 Db 396 AGTAGCATGTTGGGACGAGCGGTAGGCGACGTCTCTCGCGCGTCCAGCGGGGGA 455

QY 308 TGAGCGGCGCTGCTTGGCGCGCGTGTGTCGACACCGAGGACGAGTCCGACCGA 367  
 Db 456 GGTGATGCCGAGTAGCACTCCGAGGAGGCGTACATGTTGACACGAGCGGCGCGC 515

QY 368 TGCGCAGAGTGTGTCAGGTGGCGCTCTTCGCGCTGCGCGCTACGCCACACCGCGT 427  
 Db 516 CGCGTAGACTCCAGCGGGCGACGTACTCGCGCATGACCCCGTACGATGACGTCGA 575

QY 428 TGCCCTCGGCTTGAGCGAGTTCAGCAGGGCGGA 460  
 Db 576 TGTACTTGTGCGCGCCAGAGCGCGCGACGA 608

## RESULT 22

CG275728  
 LOCUS  
 DEFINITION  
 OGXF114TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0620D03,  
 genomic survey sequence.

CG275728  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE

CG275728.1 GI:34187869  
 GSS.

## ORGANISM

Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

1 (bases 1 to 772)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

## AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

## TITLE

Consortium for Maize Genomics

## JOURNAL

Unpublished (2002)

## COMMENT

Contact: Cathy Whitelaw

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .772

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0620D03"

/clone\_lib="ZM 0.7 1.5 KB"

/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb"

methylation filtered genomic DNA library"

[illegible][illegible]

CONTACT: Robert Strausberg, Ph.D.  
Email: [grabsr@mail.ucsf.edu](mailto:grabsr@mail.ucsf.edu)  
Address: 1000 California Avenue  
Twin Towers Building  
CDNA Library Preparation: Rubin Laboratory  
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:

```

DNA library arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2508 row: h column: 11
High quality sequence stop: 398.
Location/Qualifiers
1. .948
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6298042"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DHI0B (phage-resistant)"
/clone_lib="NIH MGC 100"
/notes="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:

```

```

/clone_id="NIH_MGC_100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:

```

EcoRI: cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

```

ORIGIN
Query Match      11.4%; Score 56.9; DB 13; Length 948;
Best Local Similarity 46.1%; Pred. No. 32;
Matches 175; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 64 CCGGCTAGACCGTGGCGCGAAGCGGTACGCGCTTCACGGTCAGCGCGCGCGGACACA 123
DB |||
QY 124 CCGCGGGGTAGCTGCTGTCACGTGACCGGGGAGACGTACAGATCCTCTCCGACGCC 193
DB |||
QY 606 GGCCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
QY 184 CGCGGGAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 243
DB |||
QY 666 CGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 725
DB |||
QY 244 CGCTAGTCCACCGCTCGCGCGTGAACCGCTGCGCGGTGCGCACACCGAGACACGTC 303
DB |||
QY 726 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 785
QY 304 GGGGTGAGCGCGCGCTCTTGGCGCGGTGCTGTCGCGCACACCGAGACACGTCGCGGA 363
DB |||
QY 786 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 845
QY 364 CCGATGCCCGCAGAAAGTTTCTGTCACGCTGCGCGCTCTTTCGCGCGCGCTACGCCAC 423
DB |||
QY 846 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 905
QY 424 CGCTGCGCTCGCGCTTACG 443
DB |||
QY 906 GNCGGCGNCCGCGCGCGCG 925

RESULT 25
AL820302/c
LOCUS AL820302 N:130 Triticum aestivum cDNA clone H01.N130.Plate_48, mRNA
DEFINITION sequence.
ACCESSION AL820302
VERSION AL820302.1 GI:21831902
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 503)
AUTHORS Wilson,I., Bewick,R., Shepherd,S., Barker,G., Parker,J., Owen,P.,
Edwards,D., Coghill,J., Holdsworth,M., Lenton,J., Shewry,P. and
Edwards,K.
A BBSRC-funded wheat EST resource for the academic community
Unpublished (2002)
Contact: Barker G
Institute of Arable Crop Research
Long Ashton, Bristol BS41 9AF United Kingdom.
FEATURES
Location/Qualifiers
1..503
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="mercia"
/db_xref="taxon:4565"
/clone="H01.N130.Plate_48"
/tissue_type="embryo"

/dev stage="2 days post germination"
/clone_lib="N:130"

ORIGIN
Query Match      11.2%; Score 55.8; DB 9; Length 503;
Best Local Similarity 51.4%; Pred. No. 43;
Matches 129; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

QY 249 GTCACACCGCGTCGGCGGTGAACGGTCTCGGGGTGCGCACCGCGGTGCGTGCCTGGGGGT 308
DB |||
QY 289 GTTCTGCGCATTTGGTATGAACAATCTGGGAATCCATCCGGAATCTCTCCAGGAGAT 230
QY 309 GAGCGCGCGCTGCTTGGCGCGCTGTCGTCGACACCGAGAGACCGAGTCCGAGACCGAT 368
DB |||
QY 229 GGTGACAGAGCTGATGCTGCTTCCGGCGCGCCACCTGCTGGCGGAGATCCCTCATCTC 170
QY 369 GCGGACAGAGTTTCTGTCAGGTGGCGCTCTTCCGCTGCGCGCTAGCGCACACCGCGCTT 428
DB |||
QY 169 CTCGAGTGTGAGACAGATGCTTCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCG 110
QY 429 GCGCTCGCGCTTGAACGAGTTTCAGCAGGCGGAGAGGTCGCTGCGGATGCGCTCGACGTC 488
DB |||
QY 109 GCGCAGCGCTTGAAGTCTGCGGTGCGACGACGTCGCGGTGCGGTGCGGTGCGGTGCG 50
QY 489 GCGCGCGGACCC 499
DB |||
QY 49 GCGCGGACCC 39

RESULT 26
CG324766
LOCUS CG324766
DEFINITION genomic survey sequence.
ACCESSION CG324766
VERSION CG324766.1 GI:34242032
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poaceae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 779)
AUTHORS Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Contact: Cathy Whitelaw
Other_GSSs: OGWBE31TH
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
Location/Qualifiers
1..779
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0526F13"
/clone_lib="ZM.0.7.1.5_KB"
/note="Vector: pBGSK; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      11.2%; Score 55.8; DB 29; Length 779;
Best Local Similarity 48.7%; Pred. No. 45;
Matches 213; Conservative 0; Mismatches 217; Indels 7; Gaps 2;

```







/clone\_lib="PTB Chimpanzee Male BAC Library"

## ORIGIN

Query Match 11.1%; Score 55.4; DB 29; Length 986;  
Best Local Similarity 44.7%; Pred. No. 53; Mismatches 247; Indels 0; Gaps 0;  
Matches 200; Conservative 0;

QY 53 CCCCCTGATGTCGGGTAGACGGTGGGCGGAGGCGGTACGGCCCTTCGACGGTTCAGCGG 112  
|||  
Db 437 CGCGCGGGGCGCGGGCGCGGCCCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496  
|||

QY 113 GCGGGCGGACACCGGCGGGTACGTGTACGTGTACGGGGGAGGTACAGATCCACT 172  
|||  
Db 497 GCG 556  
|||

QY 173 GTCCGCGAGCCCGCGCGAACTCTGCTCTTCCATGATCTGTCGGGTGTCTTCCAGG 232  
|||  
Db 557 GCCCGCCCG 616  
|||

QY 233 CGAAGACGACGCGTGTATCACCCTGCGGGCGGTGAACGCGTTCGGGGTTCGACCGGGA 292  
|||  
Db 617 CGCGGGGCG 676  
|||

QY 293 TGTGCGTTCGCGGGGTGAGCGCGGCTCTGTTGGCGGCGTGTGTCACACCCAGGAGA 352  
|||  
Db 677 CGCGGGGCGCGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736  
|||

QY 353 CAGGTTCGGAACGATGCGCGAGAAATTGTGTACGTGTGCGTTCGCGCGTTCGCGCGGT 412  
|||  
Db 737 CGCGGGGCG 796  
|||

QY 413 ACGCCACCCCGCTTGCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 472  
|||  
Db 797 CG 856  
|||

QY 473 GATGCGCTTCGACGTTCG 499  
|||  
Db 857 GCG 883  
|||

## RESULT 31

BQ080807/c  
LOCUS  
DEFINITION  
103007C04.y3 C. reinhardtii CC-1690, Deflagellation (normalized),  
Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.

ACCESSION BQ080807  
VERSION BQ080807.1 GI:22049444  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
Chlamydomonadales; Chlamydomonas.

REFERENCE  
1 (bases 1 to 456)  
Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C.,  
Lefebvre, P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.  
Analyses of the Chlamydomonas reinhardtii Genome: A Model,  
Unicellular System for Analyzing Gene Function and Regulation in  
Vascular Plants. Project: 1030

## JOURNAL

COMMENT  
Unpublished (2002)  
Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chauser@duke.edu.

Location/Qualifiers

## FEATURES

source

1..456  
/organism="Chlamydomonas reinhardtii"  
/mol\_type="mRNA"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_lib="C. reinhardtii CC-1690, Deflagellation"

(normalized), Lambda Zap II"  
/note=vector; pBluescript II SK-; Site 1: EcoRI; Site 2:  
XhoI; Deflagellation library, constructed by John Davies  
and Jeffrey McDermott, combines cDNAs from CC-1690 cells  
which had been re-synthesizing flagella for 15, 30 and 60  
min after being deflagellated by pH shock. PolyA mRNA was  
purified from each sample, pooled and cDNA synthesized.  
The cDNA was directionally cloned into lambda Zap II  
(Stratagene) in the EcoRI (5') and XhoI (3') sites.  
pBluescript II SK- plasmids were excised from the lambda  
ZAP clones by superinfection with ExAssist (Stratagene)  
phage. The library was normalized using method 4 described  
in Bernaldo et al., (1996) Genome Research 6: 791-806."

## ORIGIN

Query Match 11.0%; Score 55.2; DB 13; Length 456;  
Best Local Similarity 53.2%; Pred. No. 52;  
Matches 117; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 124 CCGCGGGTTCAGTCTGTCAGTGTACGCGGGGACGTACAGGATCCACTGTCCCGCCAGCC 183  
|||  
Db 280 CCCAGGAGGAGGTCTATGAGAGGGATGGGACCGCGTGCATGGACAGCGCGCGCGCG 221  
|||

QY 184 CGCGGGAACCTCTCTCTCTTCGCCATGATCTCGTCGCGCGTGTTCAGAGGAGAGAGCAGC 243  
|||  
Db 220 CGCGGCCAGGACGCTCGGACGCGTCTCTTGGCGAAGTCGGGTTCACAGCAGC 161  
|||

QY 244 GCGTAGTCCACCGCGTTCGGCGTGAACCGTTCGGGTTCGCGACCCGGGATGTGCGTCCG 303  
|||  
Db 160 GAGGCGAGCTCGGGTTCGGCTCGTAGCGACGCGGATCGGTCAGGAACTCGCGCGG 101  
|||

QY 304 GGGGTGAGCGCGCGCTCTTGGCGCGGCTGTGTCGCGACA 343  
|||  
Db 100 ATGATGAGCGACACCTTCCAGATCGCGCGCGCGCGCGCA 61  
|||

## RESULT 32

BQ630057/c

LOCUS

DEFINITION

HVSNB0007N02f Hordeum vulgare seedling shoot EST library

clone HVSNB0007N02f, mRNA sequence.

ACCESSION BQ630057

VERSION BQ630057.2

KEYWORDS

EST.

SOURCE

ORGANISM

Hordeum vulgare subsp. vulgare

Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Hordeum.

REFERENCE

AUTHORS

1 (bases 1 to 500)

Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D.,

Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W.,

Fenton, R.D., Oates, R. and Main, D.

Development of a genetically and physically anchored EST resource

for barley genomics: Morex drought-stressed seedling shoot cDNA

library

Unpublished (2001)

On Dec 19, 2000 this sequence version replaced gi:11894215.

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Total hg bases = 397

Seq primer: AATTAAACCTCACTAAAGG

High quality sequence stop: 497.

Location/Qualifiers

1..500

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

[illegible]

RESULT 34	
CB678003/c	
LOCUS	
DEFINITION	CB678003 736 bp mRNA linear EST 09-APR-2003 OSJUNE15I06.f OSJUNE Oryza sativa (japonica cultivar-group) cDNA clone OSJUNE15I06 5', mRNA sequence.
ACCESSION	CB678003
VERSION	CB678003.1 GI:29681728
KEYWORDS	EST.
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoidae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 736)  
 AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gta g  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 15 row: I column: 06  
 Seq primer: gta aaa cga cgg cca gta g.

FEATURES  
 source  
 1. 736  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE16N22"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEe"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

ORIGIN  
 Query Match 11.0%; Score 55; DB 14; Length 736;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAACCTCTGCTTCGCCATGATCTCGTCGGCGTGTTCACGCGAAGACGAGCGCGTA 248  
 Db 357 GAACCCCGGTTCTTGAGATGTTCCCTTGGCCAGCGTCGGGGGTACAGCCCTGAT 298  
 QY 249 GTCCACCGCGTCGGCGGTGAACCGTCCGGGGTCCGACCGGATGTGTCGGGGGT 308  
 Db 297 GGCACGCGTCGATGATGACGGCGCGCACGCGGGTCTCTCCACGCGCGGTGTGGAT 238  
 QY 309 GAGCGCGCTGCTTGGCGCGTGTGCGACACCCAGGACGAGTCCGACCGAT 368  
 Db 237 GACAGCGCACCTCGTCCATCTTTCGCTTGAACGCCCCACGCGTACGAGTCCACCGTT 178  
 QY 369 GCCGACAGAGTTCGTACGCTGGCGCTCTTGGCGTCCGCGGTACGCCACCCCGCTT 428  
 Db 177 GCTCCGCTAGATGCTCTGATGGGAGCACCCCGCTCAGCGGCTCACCACACGTTGAG 118  
 QY 429 GCCCTCGGCTTGAGCGAGTTCAGCAGGGCG 459  
 Db 117 CTGCTCCGCTTGGCGCGAGGTGCGCGCGCG 87

RESULT 35  
 CB651766/c  
 LOCUS CB651766 774 bp mRNA linear EST 08-APR-2003  
 DEFINITION OSJNE16N22.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA clone OSJNE16N22 5', mRNA sequence.  
 ACCESSION CB651766  
 VERSION CB651766.1 GI:29646759  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 774)

AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E., Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
 TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu  
 PCR Primers  
 FORWARD: gta aaa cga cgg cca gta g  
 BACKWARD: gga aac agc tat gac cat g  
 Plate: 16 row: N column: 22  
 Seq primer: gta aaa cga cgg cca gta g.

FEATURES  
 source  
 1. 774  
 Location/Qualifiers  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE16N22"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEb"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)"

ORIGIN  
 Query Match 11.0%; Score 55; DB 14; Length 774;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAACCTCTGCTTCGCCATGATCTCGTCGGCGTGTTCACGCGAAGACGAGCGCGTA 248  
 Db 701 GAACCCCGGTTCTTGAGATGTTCCCTTGGCCAGCGTCGGGGGTACAGCCCTGAT 642  
 QY 249 GTCCACCGCGTCGGCGGTGAACCGTCCGGGGTCCGACCGGATGTGTCGGGGGT 308  
 Db 641 GGCACGCGTCGATGATGACGGCGCGCACGCGGGTCTCTCCACGCGCGGTGTGGAT 582  
 QY 309 GAGCGCGCTGCTTGGCGCGTGTGCGACACCCAGGACGAGTCCGACCGAT 368  
 Db 581 GACGAGCGCACCTCGTCCATCTTTCGCTTGAACGCCCCACGCGTACGAGTCCACCGTT 522  
 QY 369 GCCGACAGAGTTCGTACGCTGGCGCTCTTGGCGTCCGCGGTACGCCACCCCGCTT 428  
 Db 521 GCTCCGCTAGATGCTCTGATGGGAGCACCCCGCTCAGCGGCTCACCACACGTTGAG 462  
 QY 429 GCCCTCGGCTTGAGCGAGTTCAGCAGGGCG 459  
 Db 461 CTGCTCCGCTTGGCGCGAGGTGCGCGCGCG 431

RESULT 36  
 CB672263/c  
 LOCUS CB672263 777 bp mRNA linear EST 09-APR-2003  
 DEFINITION OSJNEe06D11.f OSJNEb Oryza sativa (japonica cultivar-group) cDNA clone OSJNEe06D11 5', mRNA sequence.  
 ACCESSION CB672263  
 VERSION CB672263.1 GI:29675988  
 KEYWORDS EST.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
 REFERENCE 1 (bases 1 to 777)

**AUTHORS** Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
**TITLE** Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu

**PCR Primers**  
**FORWARD:** gta aaa cga cgg cca gtc  
**BACKWARD:** gga aac agc tat gac cat g  
**Plate:** 06 row: D column: 11  
**Seq primer:** gta aaa cga cgg cca gtc.

**FEATURES**  
**source**  
 1..777  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE06d11"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEa"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

**ORIGIN**  
 Query Match 11.0%; Score 55; DB 14; Length 777;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAATCTCTCTCTTGGCCATGATCTCGTCGGCGTGTTCAGCGAAGACGCGGTA 248  
 Db 703 GAACCGCGCTTCTTGGATGATGTTCCCTTGGCAGCGTTCGGCGGTACAGCGCCCTGAT 644  
 QY 249 GTCCACCGCGTGGCGGTGAACGGTTCGGGTCGCGACCGGATGTGCGTCCGGGGT 308  
 Db 643 GCGGACGCGGTGATAGCGGGCCGACCGCGGTCTCTCCACGCGCGGTGTGTGAT 584  
 QY 309 GAGCGCGCCCTCTTGGCGCGGTGTGTGTCGACACCCAGGACCGAGTCCGACCGAT 368  
 Db 583 GACGAGCGCCACCTCGTCCATCTTCGCTTGAACGCCACGCGTACGAGTCCACCCGTT 524  
 QY 369 GCGCAGAGTTCGTACGGTGGCGCTCTTGGCGTCCGCGTACGCCACCCACCCGCTT 428  
 Db 523 GCTGCGGTAGATGGTCTGCATGGGAGCACCCGCCACTCGGGGTCTACCGACACGTTGAG 464  
 QY 429 GCGCTCGGCTTGAAGCGAGTTGACGAGGCG 459  
 Db 463 CTGCTCGGCTGGCGCGAGGTGCGCGCGCG 433

**RESULT 37**  
**CB671209/c**  
**LOCUS** CB671209 810 bp mRNA linear EST 09-APR-2003  
**DEFINITION** OSJNE04K19.f OSJNEe Oryza sativa (japonica cultivar-group) cDNA clone OSJNE04K19.5, mRNA sequence.  
**ACCESSION** CB671209  
**VERSION** CB671209.1 GI:29674934  
**KEYWORDS** EST.  
**SOURCE** Oryza sativa (japonica cultivar-group)  
**ORGANISM** Oryza sativa (japonica cultivar-group)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.  
**REFERENCE** 1 (bases 1 to 810)  
**AUTHORS** Jantasuriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,

Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.  
**TITLE** Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea  
**JOURNAL** Unpublished (2003)  
**COMMENT** Contact: Rod Wing  
 Arizona Genomics Institute  
 University of Arizona  
 Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
 85721-0088, USA  
 Tel: 520 626 3967  
 Fax: 520 621 9288  
 Email: http://genome.arizona.edu

**PCR Primers**  
**FORWARD:** gta aaa cga cgg cca gtc  
**BACKWARD:** gga aac agc tat gac cat g  
**Plate:** 04 row: K column: 19  
**Seq primer:** gta aaa cga cgg cca gtc.

**FEATURES**  
**source**  
 1..810  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="mRNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /clone="OSJNE04K19"  
 /tissue\_type="Leaf"  
 /dev\_stage="3 week"  
 /lab\_host="DH10B"  
 /clone\_lib="OSJNEa"  
 /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (70-15)"

**ORIGIN**  
 Query Match 11.0%; Score 55; DB 14; Length 810;  
 Best Local Similarity 50.2%; Pred. No. 59;  
 Matches 136; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 189 GAATCTCTCTCTTGGCCATGATCTCGTCGGCGTGTTCAGCGAAGACGCGGTA 248  
 Db 690 GAACCGCGCTTCTTGGATGATGTTCCCTTGGCAGCGTTCGGCGGTACAGCGCCCTGAT 631  
 QY 249 GTCCACCGCGTGGCGGTGAACGGTTCGGGTCGCGACCGGATGTGCGTCCGGGGT 308  
 Db 630 GCGGACGCGGTGATAGCGGGCCGACCGCGGTCTCTCCACGCGCGGTGTGTGAT 571  
 QY 309 GAGCGCGCCCTCTTGGCGCGGTGTGTGTCGACACCCAGGACCGAGTCCGACCGAT 368  
 Db 570 GACGAGCGCCACCTCGTCCATCTTGGCTTGAACGCCACGCGTACGAGTCCACCCGTT 511  
 QY 369 GCGCAGAGTTCGTACGGTGGCGCTCTTGGCGTCCGCGTACGCCACCCACCCGCTT 428  
 Db 510 GCTGCGGTAGATGGTCTGCATGGGAGCACCCGCCACTCGGGGTCTACCGACACGTTGAG 451  
 QY 429 GCGCTCGGCTTGAAGCGAGTTGACGAGGCG 459  
 Db 450 CTGCTCGGCTGGCGCGAGGTGCGCGCGCG 420

**RESULT 38**  
**EX442207/c**  
**LOCUS** EX442207 924 bp mRNA linear EST 15-MAY-2003  
**DEFINITION** EX442207 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF02YD08 5-PRIME, mRNA sequence.  
**ACCESSION** EX442207  
**VERSION** EX442207.1 GI:30786042  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
**REFERENCE** 1 (bases 1 to 924)  
**AUTHORS** Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
**TITLE** Full-length cDNA libraries and normalization  
**JOURNAL** Unpublished (2001)

## COMMENT

Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF027B04QF1.

## FEATURES

Location/Qualifiers  
1. .924  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF027YD08"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

## ORIGIN

Query Match 11.0%; Score 54.8; DB 13; Length 924;  
Best Local Similarity 37.5%; Pred. No. 65;  
Matches 152; Conservative 53; Mismatches 195; Indels 5; Gaps 1;

QY 38 CGCGGAGAGAACACCCCGGTAGTCCGGGTAGACGGTGGGCGCAAGCGGTACCGCC 97  
Db 874 CGSGGGGCGNCCCGCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCC 815  
QY 98 TTGACCGGTGAGCGGGGCGGACACCGCGCGGTGAGTCCAGTGTGACCGGGGA 157  
Db 814 GSGGGGGGCGGSCCG 755  
QY 158 GTTACAGATTCAGTTC-----GCCAGCCCGCGGAACTCTCTCTCTCTCTCTCT 212  
Db 754 CGSSCGGCG 695  
QY 213 CTGTCGCGGTGTTCCAGCGCAAGAGACGCGGTGATCCACCGCGTCCGGGTGAACGC 272  
Db 694 GSGCG 635  
QY 273 GTCCGGGTGCGCACCGGATGTCGTGCGGGGTGAGCCGCGCTGTTCGCGCGCT 332  
Db 634 SCGNGCG 575  
QY 333 CGTGTGCGCACACCGAGGACAGGTCCGACCGATGCCGAGAGTTGTCACGTGCG 392  
Db 574 GGGGGGTVCCCGGCG 515  
QY 393 GCTCTTCGCGGTGCGCGGTACCGCCACCGCGCTTGCCTCGGC 437  
Db 514 CCG 470

## RESULT 39

AL538546  
LOCUS AL538546 Homo sapiens FETAL BRAIN Homo sapiens linear EST 31-MAY-2003  
DEFINITION CS0DF024YJ18 3-PRIME, mRNA sequence.

## ACCESSION

AL538546

## VERSION

1

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 1201)

## AUTHORS

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

## TITLE

Full-length cDNA libraries and normalization

## JOURNAL

Unpublished (2001)

## COMMENT

On Feb 13, 2001 this sequence version replaced gi:12802039.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 399.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DF024DE09NPI&cluster=399.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DF024DE09NPI.

## FEATURES

Location/Qualifiers  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF024YJ18"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

## ORIGIN

Query Match 11.0%; Score 54.8; DB 9; Length 1201;  
Best Local Similarity 32.2%; Pred. No. 67;  
Matches 112; Conservative 81; Mismatches 151; Indels 4; Gaps 1;

QY 152 GGGGAGCGTACAGATCCACTGTCGCGCAGCCCGCGGAATCTCTCTCTCCCATGA 211  
Db 788 GGGGAGCGAAGAAGAGMGCGGGGMMGMMGMMGMMGMMGMMGMMGMMGMMGMMG 847  
QY 212 TCTGTCGCGGTGTTCCAGCGCAAGAGCAGCGGTAGTCCACCGCGTGAACG 271  
Db 848 GMMMMGGGGGMMMMMMGGGCGMMGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 907  
QY 272 CGTCCGGGTGCGCACCGGATGTCGTGCGGGGGGTGAGCGCGCTGTTCGCGCG 331  
Db 908 GGGMMGGGGGMMMMGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 967  
QY 332 TCGTGTGCGCACACCGAGACAGGTCCGACCGATCCCGCAGAGTTCTCTACGTTG 391  
Db 968 KCCGCGKCCCG 1027  
QY 392 CGCTTTCGCGGTGCGCGGTACCGCACCGCGCTTGCCTCGGCTTGAGCGAGTTCA 451  
Db 1028 SGC-----CCCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1083  
QY 452 CGAGGGGAGCAGTCCGTGCGGATGCCCTCGAGTCGCGCGCGCAACC 499  
Db 1084 GGGGGGGGCG 1131

## RESULT 40

CA075606/c  
LOCUS SCJLAMI062B06, g AM1 Saccharum officinarum cDNA clone SCJLAMI062B06  
DEFINITION 5', mRNA sequence.

## ACCESSION

CA075606

## VERSION

EST.

## KEYWORDS

EST.

## SOURCE

Saccharum officinarum

## ORGANISM

Saccharum officinarum

## REFERENCE

1 (bases 1 to 605)

## AUTHORS

Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.



```

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM0626G15"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: PCR1-TOPO; Site: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

FEATURES
source
1..905
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM0626G15"
/clone_lib="ZM 0.6 1.0 KB"
/notes="Vector: PCR1-TOPO; Site: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

ORIGIN
Query Match 10.8%; Score 54.2; DB 28; Length 905;
Best Local Similarity 48.8%; Pred. No. 79;
Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 199 TCCTTCGCCATGATCGTCGGCGTGTCCAGGCGAAGAGCAGCGCGTAGTCCACCGG 258
Db 439 TCGGCTTCCTTGACAGCGCAGCGCGGCCCTCGCGCCCGTCACTTGAGCAGCAG 498
QY 259 TCGGGCGTGAAACCGCTCCGGGTGCGACCGGGATGTGCGTCCGGGGTGAGCCGCCCC 318
Db 499 TCCCGCTTGGCACCCTTGCGGTGAGCGGCTTGAGGTGCTTACGCGGGCCCTTCCGGTCC 558
QY 319 TGCTTGGCGGCGTGTGTCGACACCCAGAGCAGGTCCGACCGATGCGGACAGAG 378
Db 559 ACGGGGAACACCTTGAAGACCCCGGACCATGAGTGCACCGCGCGTTCAGGGCCCTGGACG 618
QY 379 TTCGTACAGTGGCGCTTTCGCCGTGCGCGGTACGCCACACCGCGTTCGCCCTCGGCC 438
Db 619 TACACCATGACGCGGTGCGCGCTTTCGCCGTGCGCGGTGCGCGGTGCGCGGTGCGTAC 678
QY 439 TTGAGCAGTTTACAGAGCGGACGAGTGTGCGATGCGTCCGATGCGTCCGCGCGGAA 497
Db 679 CAGCGTGTGCGCGCGCCAGACCTGCGCGCGCTTGATCCACTGCTGTCGACGCGGAA 737

RESULT 43
CG268544/c 916 bp DNA linear GSS 25-AUG-2003
LOCUS CGX444TV.ZM.0.7.1.5.KB.Zea mays genomic clone ZMMBMA0626G16,
DEFINITION genomic survey sequence.
ACCESSION CG268544
VERSION CG268544.1 GI:34180685
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 916)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: CGX444TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
Location/Qualifiers

FEATURES
source
1..916
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBMA0626G16"
/clone_lib="ZM 0.7 1.5 KB"
/notes="Vector: pBCSk-; Site: 1; HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 10.8%; Score 54.2; DB 29; Length 916;
Best Local Similarity 48.8%; Pred. No. 80;
Matches 146; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 199 TCCTTCGCCATGATCGTCGGCGTGTTCACGCGAAGAGCAGCGGTATCCACCGG 258
Db 384 TCGGCTTCCTTGACAGCGCAGCGCGGCCCTCGCGCCCGTCACTTGAGCAGCAGC 325
QY 259 TCGGGCGTGAAACCGCTCCGGGTGCGACCGGGATGTGCGTCCGGGGTGAGCCGCCCC 318
Db 324 TCCCGCTTGGCACCCTTGTGCGTGTGAGCGCTTGAGGTGCTTGACGGGGCTTCGGTCC 265
QY 319 TGCTTGGCGGCGTGTGTCGACACCCAGAGCAGGTCCGACCGATGCGGACAGAG 378
Db 264 ACGGGGAACACCTTGAAGACCCCGGACCATGAGTTCGACCGCGCGTTCAGGGCCCTGGACG 205
QY 379 TTCGTACAGTGGCGCTTTCGCCGTGCGCGGTACGCCACACCGCTTTCGCCCTCGGCC 438
Db 204 TACACCATGACGCGGTGCGCGCTTTCGCCGTCCACGCGGAGCGCGCCACCTCGTAC 145
QY 439 TTGAGCAGTTTACAGAGCGGCGAGCAGTGTGCGTCCGGATGCGTCCGAGTCCGCGGAA 497
Db 144 CAGCGCTGTGCGCGCGCAGGACCTGCGCGCGTTCATCCACTGCTGTCGACGCGGAA 86

RESULT 44
CG258758/c 596 bp mRNA linear EST 27-NOV-2002
LOCUS BR030004000.PLATE_E01_5_005.ab1.OA.Oryza sativa (japonica
DEFINITION cultivar-group) cDNA clone BR030004000.PLATE_E01_5_005.ab1 similar
to contains ESTs AU097436 (S4951), D41936 (S4951), AU101944 (S2163),
AU095016 (E50186), AU058216 (E50186), C72995 (E2631),
D40291 (S2163), D41387 (S3866) unknown protein [Oryza sativa (japo,
mRNA sequence.
CA758758
CA758758.1 GI:25802797
EST.
ORyza sativa (japonica cultivar-group)
ORyza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrentoidae; Oryzae; Oryza.
1 (bases 1 to 596)
Bohnert, H.J., Borchert, C., Brazille, S., Brooks, J., Eaton, M.,
Ferreira, H., Kawasaki, S., McColough, A., Michalowski, C.B.,
Palacio, C., Scara, G., Wheeler, M. and Zepeda, G.R.
Functional Genomics of Plant Stress Tolerance
Unpublished (2000)
Contact: Mark Fredrickson
Department of Plant Biology
University of Illinois
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 2172655473
Email: bohnertlab@life.uiuc.edu
Location/Qualifiers
1..596
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="BR030004000.PLATE_E01_5_005.ab1"
/tissue_type="roots"

FEATURES
source
1..596
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/db_xref="taxon:39947"
/clone="BR030004000.PLATE_E01_5_005.ab1"
/tissue_type="roots"

```



```

/dev_stages="3-4 weeks"
/clone_lib="OA"
/note="19 h 200mM NaCl"

ORIGIN
Query Match      10.8%; Score 54; DB 14; Length 596;
Best Local Similarity 49.8%; Pred. No. 81;
Matches 135; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 189 GAACCTCTCTCTTCGCGCATGATCTCGTCGGCGTGTTCAGGCGAAGACGAGCGGTA 248
DB 366 GAACCCGCGCTCTCTNAGCATGTTCCCTTGGCCAGCTCGCGGGTACAGCGCCCTGAT 307

QY 249 GTTCACCGCGTTCGGCGGTGAACGCTCCGCGGGTGGCACCGGGATGTGCGTGGCGGGGT 308
DB 306 GCGCAGCGCGCTGATGAGCGGCGCCGACGCGCGGTCTCTCTCCACGCGCGGGTTGTGGAT 247

QY 309 GAGCGCGCGCTCTTGGCGGGTGTGTCGTGCGACACCCAGGAGACCGAGTCCGACCGAT 368
DB 246 GACGAGCGCCACCTGCTCATCTTTCGCTTGAACGCCCCACCGTACGAGTCCCAACCGTT 187

QY 369 GCGCGAGAGTGTGTCAGCGTGGCGCTCTTCGCGGTTCGCGGTTCGCGGTACGCCACCCCGCTT 428
DB 186 GCTGCGGTAGATGCTGTCATGGGAGCACCCCGCTCGCGGGTCAACCGACAGTTGAG 127

QY 429 GCGCTCGCGCTTACCGAGTTCAGCAGGCG 459
DB 126 CTGCTCCGCGTGGCGAGTTCGCGCGCG 96

RESULT 45
LOCUS CG447961 866 bp DNA linear GSS 17-SEP-2003
DEFINITION OG9AA78TH_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZM58A0849M11,
Genomic survey sequence.
ACCESSION CG447961
VERSION CG447961.1 GI:34832961
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 866)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
JOURNAL
COMMENT Other_GSSs: OG9AA78TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..866
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM58A0849M11"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      10.8%; Score 54; DB 29; Length 866;
Best Local Similarity 49.0%; Pred. No. 85;
Matches 144; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 189 GAACCTCTCTCTTCGCGCATGATCTCGTCGGCGTGTTCAGGCGAAGACGAGCGGTA 248
DB 366 GAACCCGCGCTCTCTNAGCATGTTCCCTTGGCCAGCTCGCGGGTACAGCGCCCTGAT 307

QY 249 GTTCACCGCGTTCGGCGGTGAACGCTCCGCGGGTGGCACCGGGATGTGCGTGGCGGGGT 308
DB 306 GCGCAGCGCGCTGATGAGCGGCGCCGACGCGCGGTCTCTCTCCACGCGCGGGTTGTGGAT 247

QY 309 GAGCGCGCGCTCTTGGCGGGTGTGTCGTGCGACACCCAGGAGACCGAGTCCGACCGAT 368
DB 246 GACGAGCGCCACCTGCTCATCTTTCGCTTGAACGCCCCACCGTACGAGTCCCAACCGTT 187

QY 369 GCGCGAGAGTGTGTCAGCGTGGCGCTCTTCGCGGTTCGCGGTTCGCGGTACGCCACCCCGCTT 428
DB 186 GCTGCGGTAGATGCTGTCATGGGAGCACCCCGCTCGCGGGTCAACCGACAGTTGAG 127

QY 429 GCGCTCGCGCTTACCGAGTTCAGCAGGCG 459
DB 126 CTGCTCCGCGTGGCGAGTTCGCGCGCG 96

RESULT 46
LOCUS CG338271 851 bp DNA linear GSS 26-AUG-2003
DEFINITION OG0CG91TH_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZM58Ma0685P13,
Genomic survey sequence.
ACCESSION CG338271
VERSION CG338271.1 GI:34255537
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 851)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
JOURNAL
COMMENT Other_GSSs: OG0CG91TV
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
source
1..851
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM58Ma0685P13"
/clone_lib="ZM 0.7 1.5 KB"
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match      10.8%; Score 53.8; DB 29; Length 851;
Best Local Similarity 49.6%; Pred. No. 91;
Matches 170; Conservative 0; Mismatches 167; Indels 6; Gaps 1;

QY 109 GCGCGCGCGCGACACACCGCGGTTCAGTCTGTCAGTGTACCGCGGGACGTACAGGATC 168
DB 23 GCGCTCCCGCCACCTCCGACCGCGGTGACCTCATCTGCACGCCCGCGACGTGAGGCTT 82

QY 169 CACTGTCCCGCAGCCCGCGGAACTCTCTGCTCTTCGCGCATGATCTCTGCGCGGTTC 228
DB 83 CGTACCTTGAACCTCCCGCGGTGTAGCCCGCGGCCCAACAGCTCGCGACTGCACG 142

```

```

QY 229 CAGGCGAAGACGAGCGGTAGTCCACCGCGTGGGGGTGAACGCGTCCGGGTGGCGACC 288
Db 143 ACGGCGAACATGCGGCGCGCCACCGCTCTGATACCGCGCGACGGTC-----GGCGACC 196
QY 289 GGGATGCGGTGCGCGGGGGTGAACCGCCCTGCTTGGCCCGCGTCTGTGCGACACCCAG 348
Db 197 TGGATGCGCGCCACACCGAGGTGACGCGCCCTGCGAGGCGACAGACTCGGACCACTCGTCC 256
QY 349 GAGACGAGTCCGCGGACCGATGCGCAGAGATTCGTACGTTGGCGGTCTTCCCGGTGCGG 408
Db 257 GAGGGACACGACGAGTGCAGACCGCCGTGTCGCCCGCGCGCGGAGACCTCGAACGGC 316
QY 409 CCGTACGCCACACCGCCCTTGCCTCGGCTTGAAGCGAGTTCA 451
Db 317 GCGTCTGTCGCGCCATCTCGGCGTGGACGCGCGCGCGTCCA 359

RESULT 47
AG171124/c
LOCUS Pan troglodytes DNA, clone: RP43-040F09.TJ, genomic survey
DEFINITION
ACCESSION AG171124.1 GI:16700802
VERSION
KEYWORDS
SOURCE
ORGANISM
Pan troglodytes (chimpanzee)
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPCI-43
Unpublished
REFERENCE
2 (bases 1 to 1798)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
Location/Qualifiers
1. 1798
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-040F09.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"

ORIGIN
Query Match 10.8%; Score 53.8; DB 29; Length 1798;
Best Local Similarity 42.9%; Pred. No. 99;
Matches 186; Conservative 0; Mismatches 247; Indels 1; Gaps 1;

QY 40 GCAGAGGAGAACCCCGGTATCGGGGTAGACGGTGGCGCGGAAGCGGTACGGCCCTT 99
Db 1134 GNGNGGGCGGCCCGCGNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075
QY 100 CGACGGGTACGCGCGCGCGCGACACCGCGCGGTACGTCGTCTACGCGGTACGCGGGGACG 159

```

```

Db 1074 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1015
QY 160 TACAGGATCCACTGTCGCGCAGCCCGCGCGGAACTCTCTGCTCTTCGCGCATGATCTCGTCG 219
Db 1014 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955
QY 220 GCGTGGTTCCAGGCGAAGAGCAGCGCGTAGTCCACCGCGTGGCGCGTGAACGCGTCCCGG 279
Db 954 GCGNNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 896
QY 280 GTGCGCACCGGGATGTCGTCGCGCGGGGTGAGCGCGCCCTGCTGTCGCGCGCGCGTCTGTCG 339
Db 895 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 836
QY 340 CACACCCAGGAGACAGGTCGCGGACCGATGCGCGCAGAAATTCGTCACGCGTGGCGCTTTC 399
Db 835 GCGNNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 776
QY 400 GCCTGCGCGCGGTACGCCACCGCTTGCCTTCGCGCTTGAGCGAGTTTCAGCAGGGCG 459
Db 775 NCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 716
QY 460 AGCAGTCCGTGCG 473
Db 715 CGCGCGCGCGCGCG 702

CB659824
OSNED01316.f OSJNED Oryza sativa (japonica cultivar-group) cDNA
clone OSNED01316 5', mRNA sequence.
ACCESSION CB659824
VERSION
KEYWORDS
SOURCE
EST.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Embryophyta; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Oryza.
REFERENCE
1 (bases 1 to 665)
Jantaauriyarat, C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E.,
Kudrna, D., Dean, R., Soderlund, C., Wing, R. and Wang, G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 488A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 01 row: B column: 15
Seq primer: gta aac cga cgg cca gtc.
Location/Qualifiers
1. 665
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSNED01316"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNED"
/note="Vector: pBLUEScript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"
FEATURES
source

```



```

Db      370 GCCGTAACGATGCTCTCTGCTGGATTAAATCACACGCTGTAGATACTCTCGAAGCGCGACG 311
QY      245 CGTAGTCCACCGGTCGGGGCGTGAAACGCGTCCGGGGTCCGCACCGGGATGTGGTGGCCGG 304
Db      310 ACACACACGGGTGATGGAGGCGGTGGACCCGGAGGTGTGGGTGACGTGCACGGACATGAACC 251
QY      305 GGGTGAACCGCGCCCTGCTTGGCCGGCGTCTGTGCGCACACCCAGGAGACCCAGGTCCGGAC 364
Db      250 TGGTGGGAGAGGCTTCCAGCTCGCGCTGCTGTGCACCCAGCGCCACCCCGGGACCGGC 191
QY      365 CGATGCCGCGAGAGTTCTGTACGGTGGCGCTCTTGGCCGTCGGCGGTAGCCACACCC 424
Db      190 CGACCATGCAACGAGGTGCGAGGGTGTCTCTCCCTCTCTGCGCGCCCGCGCAAGCCTC 131
QY      425 GCTTGGCCCTCGGCCTTGAGCGAGTTACAGAGGCGAGCGAGGTGGGTGGGTGGGTGGGT 484
Db      130 CCGGCTCCAGGCGGCGCGGCTTTCGGCGGCGCGCGGCTTACACGCGCTTCTCGCGA 71
QY      485 CGTGGCGG 493
Db      70 CCGCGGCGG 62

```

Search completed: June 27, 2004, 20:50:43  
 Job time : 2035.43 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:09 ; Search time 325.98 Seconds  
(without alignments)  
6516.033 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_1\_500  
Perfect score: 500  
Sequence: 1 ggtaccgaccgtgtccgg.....tcgacgtcggcggaacct 500

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues  
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001as:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	500	100.0	37116	7	ABZ66810	Abz66810 Orthosomy
2	497.2	99.8	13519	5	ABZ66785	Abz66785 Orthosomy
3	369	73.8	1251	7	ABZ66785	Abz66785 Orthosomy
4	298.4	59.7	10035	7	ABZ66813	Abz66813 Orthosomy
5	298.4	59.7	11115	4	ABL50562	ABL50562 Micromono
6	256.6	51.3	1251	7	ABZ66784	Abz66784 Orthosomy
7	149.8	30.0	1401	2	AAV34002	AAV34002 S. peucet
8	135	27.0	16020	3	AAA3283	Aaa3283 Streptomy
9	131	26.2	12152	5	AA508699	AA508699 M. carbon
10	124	24.8	603	7	ABZ66779	Abz66779 Orthosomy
11	118.2	23.6	15240	7	ABZ66809	Abz66809 Orthosomy
12	112.8	22.6	1281	7	ABZ66730	Abz66730 Orthosomy
13	97	19.4	42000	3	AAA63349	Aaa63349 Streptomy
14	97	19.4	63164	3	AAA63348	Aaa63348 Streptomy
15	95	19.0	25617	8	ACC58256	Acc58256 Novobioc
16	94	18.8	31248	8	ACC58250	Acc58250 Coumermyc
17	94	18.5	35359	8	ACC58251	Acc58251 Coumermyc
18	92.4	18.5	1290	7	ABZ66729	Abz66729 Orthosomy
19	92.4	18.5	14252	4	ABL50558	ABL50558 Micromono
20	92.4	18.5	48221	7	ABZ66811	Abz66811 Orthosomy
21	90.6	18.1	45055	7	ABZ66808	Abz66808 Orthosomy
22	87.8	17.6	1224	7	ABZ66728	Abz66728 Orthosomy
23	81	16.2	59616	7	ABZ37516	Abz37516 Streptomy

C 24	81	15.2	59816	7	ABZ37515	Abz37515 Streptomy
C 25	80.8	15.2	1224	7	ABZ37534	Abz37534 Streptomy
C 26	78	15.6	42291	8	ACC58253	ACC58253 Clorobloc
C 27	59.4	11.9	624	7	ABZ66778	Abz66778 Orthosomy
C 28	56.8	11.4	2122	2	AA70152	Aat70152 S. longisp
C 29	55.2	11.0	9994	4	AA85191	Aac85191 S. avermi
C 30	54.8	11.0	837	7	ACA27222	ACA27222 Prokaryot
C 31	54.6	10.9	9975	7	AA161173	AA161173 Actinosyn
C 32	54.6	10.9	8746	7	AA161224	AA161224 Actinosyn
C 33	54.4	10.9	3510	3	AA65259	AA65259 Maize Sca
C 34	53.6	10.7	2964	7	ADA70431	Ada70431 Rice gene
C 35	53.4	10.7	1551	7	ACA37987	ACA37987 Prokaryot
C 36	53.4	10.7	27541	4	AA17185	AA17185 Streptomy
C 37	53.4	10.7	125401	4	AA17186	AA17186 Streptomy
C 38	53.2	10.6	2700	4	AA18767	AA18767 Human ANI
C 39	52.2	10.4	1200	7	ACA37556	ACA37556 Prokaryot
C 40	51.4	10.3	2742	7	AA161170	AA161170 Actinosyn
C 41	51.2	10.2	975	7	ABZ66676	ABZ66676 Orthosomy
C 42	51.2	10.2	2010	7	ACA23382	ACA23382 Prokaryot
C 43	50.6	10.1	546	2	AA76906	Aat76906 S. glauce
C 44	50.6	10.1	2697	7	ADA70563	Ada70563 Rice gene
C 45	50.6	10.1	88421	6	AA140781	AA140781 88421nt g
C 46	50.4	10.1	23666	2	AAQ10190	AAQ10190 Cephalosp
C 47	50.4	10.0	1248	7	ACA37895	ACA37895 Prokaryot
C 48	50.2	10.0	1605	2	AA02914	Aax02914 Z. mays c
C 49	50.2	10.0	6733	2	AA02913	Aax02913 Z. mays c
C 50	50.2	10.0	6733	2	AA02913	Aax02913 Z. mays c

## ALIGNMENTS

RESULT 1  
ABZ66810/c  
ID ABZ66810 standard; DNA; 37116 BP.

XX AC ABZ66810;  
XX DT 27-OCT-2003 (revised)  
XX DT 21-MAR-2003 (first entry)

XX Orthosomycin biosynthetic gene cluster SEQ ID NO 279.

XX Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.

XX Micromonospora carbonacea; africana.

XX WO200279505-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-CA000432.

XX 28-MAR-2001; 2001US-0279095P.

XX 30-MAR-2001; 2001US-027909P.

XX 20-APR-2001; 2001US-0285214P.

XX (ECOP-) ECOPIA BIOSCIENCES INC.

XX Farnet CM, Zazopoulos E, Staiffa A;

XX WPI; 2003-0584435/05.

XX Identifying orthosomycin biosynthetic gene, gene fragment or gene cluster, by detecting presence of nucleic acid sequence corresponding to 17 of flammulicins protein families.

XX Example 10; Page 458-478; 511pp; English.

XX The invention relates to identifying orthosomycin biosynthetic genes and its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the presence of a nucleic acid sequence coding for a polypeptide (ABP99207-ABP99362). The method is useful for identifying an orthosomycin

CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 27-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 37116 BP; 5171 A; 13140 C; 13619 G; 5186 T; 0 U; 0 Other;  
Query Match 100.0%; Score 500; DB 7; Length 37116;  
Best Local Similarity 100.0%; Pred. No. 9e-83;  
Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGTACCCGACCGTCTCCGCGACACGAGTCCAGATACGCGGAGGACACCCCGCGT 60  
Db 30486 GGTACCCGACCGTCTCCGCGACACGAGTCCAGATACGCGGAGGACACCCCGCGT 30427  
QY 61 AGTCGCGGTAGACGGTGGCGGCGGAGGCGGTACGCGCTTCGACGCGTCCGCGGCGG 120  
Db 30426 AGTCGCGGTAGACGGTGGCGGCGGAGGCGGTACGCGCTTCGACGCGTCCGCGGCGG 30367  
QY 121 ACACCGGGGGTACAGTGTACGCTGTACGCGGGGAGGACGATCCACTGTCCGCCA 180  
Db 30366 ACACCGGGGGTACAGTGTACGCTGTACGCGGGGAGGACGATCCACTGTCCGCCA 30307  
QY 181 GCGCGGGCGAACTCCTGTCTCTTCCCATGATCTCGTCGGGTGTTCCAGCGCAAGAGC 240  
Db 30306 GCGCGGGCGAACTCCTGTCTCTTCCCATGATCTCGTCGGGTGTTCCAGCGCAAGAGC 30247  
QY 241 AGCGGTGTAGTCACCGGTGCGGGCGTGAACGGGTCCGGGTGCGGACCGGGATGTGGTG 300  
Db 30246 AGCGGTGTAGTCACCGGTGCGGGCGTGAACGGGTCCGGGTGCGGACCGGGATGTGGTG 30187  
QY 301 CCGCGGGGTAGCGCGGCGCTGTGTCGCGGCGTGTGCGCACCCAGGACACCAAGTCC 360  
Db 30186 CCGCGGGGTAGCGCGGCGCTGTGTCGCGGCGTGTGCGCACCCAGGACACCAAGTCC 30127  
QY 361 GAACCGATCCCGAGAGTTCGTACGGTTCGGGTCTTTCGCGCGTCCGCGCGTACGCCACC 420  
Db 30126 GAACCGATCCCGAGAGTTCGTACGGTTCGGGTCTTTCGCGCGTCCGCGCGTACGCCACC 30067  
QY 421 ACCCGCTTCCCTCGGCTTGAAGGATTCAGCAGCGGAGCAGGTCCGTCGGATGCC 480  
Db 30066 ACCCGCTTCCCTCGGCTTGAAGGATTCAGCAGCGGAGCAGGTCCGTCGGATGCC 30007  
QY 481 TCGACGTCCGCGCGAACT 500  
Db 30006 TCGACGTCCGCGCGAACT 29987  
RESULT 2  
AAS08693  
ID AAS08693 standard; DNA; 109519 BP.  
XX AC  
XX AAS08693;  
XX  
XX 14-SEP-2003 (revised)  
DT 25-SEP-2003 (first entry)  
XX  
XX DE Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.  
XX Everninomycin; antibiotic; bottle-neck gene; orthomycin; fermentation;  
XX ds.  
XX  
XX Micromonospora sp. ATCC 39149.  
XX  
XX Key Location/Qualifiers  
FH complement(132..1382)  
FT CDS  
FT /\*tag= a

FT RBS  
FT complement(1389..1394)  
FT /\*tag= b  
FT CDS complement(1490..2611)  
FT /\*tag= c  
FT /\*product= "EvdB"  
FT complement(2618..2622)  
FT /\*tag= d  
FT CDS complement(2622..3860)  
FT /\*tag= e  
FT /\*product= "EvdC"  
FT complement(3867..3870)  
FT /\*tag= f  
FT RBS 4134..4138  
FT /\*tag= h  
FT CDS 4143..5312  
FT /\*tag= g  
FT /\*product= "EvdD"  
FT 5309..6235  
FT /\*tag= i  
FT /\*product= "EvdE"  
FT 6226..6229  
FT /\*tag= k  
FT CDS 6232..7275  
FT /\*tag= j  
FT /\*product= "EvdF"  
FT 7272..8527  
FT /\*tag= l  
FT /\*product= "EvdG"  
FT 8333..8336  
FT /\*tag= n  
FT CDS 8342..9364  
FT /\*tag= m  
FT /\*product= "EvdH"  
FT complement(9463..10224)  
FT /\*tag= o  
FT RBS /\*product= "EvdI"  
FT complement(10232..10235)  
FT /\*tag= p  
FT CDS 10424..11176  
FT /\*tag= q  
FT /\*product= "EvdJ"  
FT 12027..12455  
FT /\*tag= r  
FT CDS /\*product= "EvdK"  
FT /\*note= "No start codon"  
FT complement(12108..13022)  
FT /\*tag= B  
FT RBS /\*product= "EvdL"  
FT complement(13027..13030)  
FT /\*tag= t  
FT CDS complement(14410..15363)  
FT /\*tag= u  
FT /\*product= "EvrA"  
FT complement(15369..15373)  
FT /\*tag= v  
FT CDS complement(15380..16414)  
FT /\*tag= w  
FT /\*product= "EvrB"  
FT complement(16419..17873)  
FT /\*tag= x  
FT /\*product= "EvrC"  
FT complement(17870..18934)  
FT /\*tag= y  
FT RBS /\*product= "EvrD"  
FT 19374..20906  
FT /\*tag= z  
FT /\*product= "EvrE"  
FT 21056..22542  
FT /\*tag= ab  
FT CDS 21064..22542  
FT /\*tag= aa

```
FT RBS /product= "EvrF" 22736. .22740
FT FT /*tag= ad 22748. .24172
FT CDS /*tag= ac 22748. .24172
FT CDS /product= "EvrG" complement(24177. .25223)
FT FT /*tag= ae 22748. .24172
FT RBS /product= "EvrH" complement(25230. .25233)
FT FT /*tag= af 22748. .24172
FT CDS 25550. .26626
FT FT /*tag= ag 22748. .24172
FT RBS /product= "EvrI" 26672. .26676
FT FT /*tag= ai 26685. .30479
FT CDS /*tag= ah 26685. .30479
FT FT /product= "EvrJ" complement(30557. .31876)
FT CDS /*tag= aj 26685. .30479
FT FT /product= "EvrK" complement(31885. .31888)
FT RBS /*tag= ak 26685. .30479
FT CDS complement(31941. .32882)
FT FT /*tag= al 26685. .30479
FT CDS /product= "EvrL" complement(33167. .34405)
FT FT /*tag= am 26685. .30479
FT RBS /product= "EvrM" complement(34414. .34418)
FT FT /*tag= an 26685. .30479
FT CDS complement(34449. .35210)
FT FT /*tag= ao 26685. .30479
FT RBS /product= "EvrN" complement(35219. .35221)
FT CDS /*tag= ap 26685. .30479
FT FT complement(35294. .36238)
FT CDS /*tag= aq 26685. .30479
FT CDS /product= "EvrO" complement(36235. .36963)
FT FT /*tag= ar 26685. .30479
FT CDS /product= "EvrP" complement(36998. .38026)
FT FT /*tag= as 26685. .30479
FT CDS /product= "EvrQ" complement(38072. .38566)
FT FT /*tag= at 26685. .30479
FT CDS /product= "EvrR" complement(38892. .40163)
FT FT /*tag= au 26685. .30479
FT CDS /product= "EvrS" complement(40216. .40890)
FT FT /*tag= av 26685. .30479
FT CDS /product= "EvrT" complement(40887. .41576)
FT FT /*tag= ax 26685. .30479
FT RBS /product= "EvrU" complement(40899. .40902)
FT FT /*tag= aw 26685. .30479
FT CDS complement(41679. .42707)
FT FT /*tag= ay 26685. .30479
FT RBS /product= "EvrV" complement(42714. .42717)
FT FT /*tag= az 26685. .30479
FT CDS complement(42810. .43799)
FT FT /*tag= ba 26685. .30479
FT CDS /product= "EvrW" complement(43799. .44866)
FT FT /*tag= bc 26685. .30479
FT RBS /product= "EvrX" complement(43807. .43811)
FT FT /*tag= bb 26685. .30479

complement(45014. .45760)
/*tag= bd
/product= "EvrY"
complement(45767. .45770)
/*tag= be
complement(45952. .45956)
/*tag= bg
complement(45962. .46714)
/*tag= bf
/product= "EvrZ"
complement(47156. .49234)
/*tag= bh
/product= "EvrA"
51627. .52715
/*tag= bi
/product= "EvrB"
51629. .51622
/*tag= bj
52889. .53557
/*tag= bk
/product= "EvrC"
53554. .54207
/*tag= bl
/product= "EvrA"
complement(54362. .55117)
/*tag= bm
/product= "EvrB"
complement(55125. .55128)
/*tag= bn
complement(55135. .56094)
/*tag= bo
/product= "EvrC"
complement(56100. .56103)
/*tag= bp
complement(56184. .56813)
/*tag= bq
/product= "EvrC2"

Query Match 99.8%; Score 499.2; DB 5; Length 109519;
Best Local Similarity 99.6%; Pred. No. 1.1e-82;
Matches 498; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTACCCGACCGTGTCCCGGACACGAGTGTGAGTACGGGAGAGGACACCCCGGT 60
DB 1 GGTACCCGACCGTGTCCCGGACACGAGTGTGAGTACGGGAGAGGACACCCCGGT 60
QY 61 AGTCCGCGTAGACCGTGTCCCGGACACGAGTGTGAGTACGGGAGAGGACACCCCGGT 120
DB 61 AGTCCGCGTAGACCGTGTCCCGGACACGAGTGTGAGTACGGGAGAGGACACCCCGGT 120
QY 121 ACACCGCGCGGTACGCTGTACGTTACGCGGGGACGTACAGGATCCTGTCGCGCA 180
DB 121 ACACCGCGCGGTACGCTGTACGTTACGCGGGGACGTACAGGATCCTGTCGCGCA 180
QY 181 GCCCGGCGGAACCTCTCTCTTCCCATGATCTGTCGGGTGTTCCAGCGAAGAGC 240
DB 181 GCCCGGCGGAACCTCTCTCTTCCCATGATCTGTCGGGTGTTCCAGCGAAGAGC 240
QY 241 AGCGGTAGTCCACCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 300
DB 241 AGCGGTAGTCCACCGCGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 300
QY 301 CCGGGGTTGAGCGCGGCGCTTGTGTCGCAACCGCGTGTGTCGCAACCGCGGAGTGTG 360
DB 301 CCGGGGTTGAGCGCGGCGCTTGTGTCGCAACCGCGTGTGTCGCAACCGCGGAGTGTG 360
QY 361 GGACCGATGCCGACAGAGTTCGTACGTTGCGGTCTTTTCGCGGTGCGCGGTGCGCGTACC 420
DB 361 GGACCGATGCCGACAGAGTTCGTACGTTGCGGTCTTTTCGCGGTGCGCGGTGCGCGTACC 420
QY 421 ACCCGTTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 480
DB 421 ACCCGTTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 480
```



```
QY 481 TCGACGTGCGCGGCAACT 500
DB 481 TCGACGTGCGCGGCAACT 500

RESULT 3
ABZ66785/c
ID ABZ66785 standard; DNA; 1251 BP.
XX
AC ABZ66785;
XX
DT 27-OCT-2003 (revised)
DT 21-MAR-2003 (first entry)
XX
DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 232.
XX
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Micromonospora carbonacea; africana.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
PR 30-MAR-2001; 2001US-0279709P.
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Zazopoulos E, Staffa A;
XX
DR WPI; 2003-058435/05.
DR P-FSDB; ABP99322.
XX
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
PT cluster, by detecting presence of nucleic acid sequence corresponding to
PT 17 of flambamycins protein families.
XX
PS Claim 16; Page 382; 51lpp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
CC ABP99362). The method is useful for identifying an orthosomycin
CC biosynthetic gene, gene fragment or gene cluster, especially an
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
CC gene fragment or gene cluster. The method is useful for detecting the
CC presence of any organism that contains DNA for the production of
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
CC orthosomycins) regardless of the level at which genes for orthosomycin
CC production are expressed by the organism or the amount of orthosomycin
CC produced by the organism. This allows for the detection of new
CC orthosomycin natural products, not produced by the organism. (Updated on
CC 27-OCT-2003 to standardise OS field)
XX
SQ Sequence 1251 BP; 181 A; 438 C; 439 G; 193 T; 0 U; 0 Other;
XX
Query Match 73.8%; Score 369; DB 7; Length 1251;
Best Local Similarity 100.0%; Pred No. 9.6e-59;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 132 TCGAGTGTGTCAGCGGGGACGTACAGATCCATGTCGCCAGCGCGGAA 191
DB 1251 TCGAGTGTGTCAGCGGGGACGTACAGATCCATGTCGCCAGCGCGGAA 1192
QY 192 CTCTGCTCTTCCCATGATCTCGTCGGCGTTCAGGGCGAAGACGACGCGTAGNC 251
DB 1191 CTCTGCTCTTCCCATGATCTCGTCGGCGTTCAGGGCGAAGACGACGCGTAGTC 1132
```

```
QY 252 CACCGCGTCGGCGGTGAAACGCGTCCGGGFTGCCACCGGATGTGCGTCCCGGGGTGAG 311
DB 1131 CACCGCGTCGGCGGTGAAACGCGTCCGGGFTGCCACCGGATGTGCGTCCCGGGGTGAG 1072
QY 312 CCGGCCCTGCTTGGCGGGTGTGTCGCACACCCACAGGAGACGAGTCCGAGCCGATGCC 371
DB 1071 CCGGCCCTGCTTGGCGGGTGTGTCGCACACCCACAGGAGACGAGTCCGAGCCGATGCC 1012
QY 372 GCAGAGATTCTGTACCGTGGCGCTCTTTCGCCGTCGCGCGGTACGCCACACCGCTTGCC 431
DB 1011 GCAGAGATTCTGTACCGTGGCGCTCTTTCGCCGTCGCGCGGTACGCCACACCGCTTGCC 952
QY 432 CTCGGCCTTGAGCGAGTTTCAGCAGGGGAGCAGGTGCGGATGCCCTTCGACGTCGGC 491
DB 951 CTCGGCCTTGAGCGAGTTTCAGCAGGGGAGCAGGTGCGGATGCCCTTCGACGTCGGC 892
QY 492 GCGGAACCT 500
DB 891 GCGGAACCT 883

RESULT 4
ABZ66813/c
ID ABZ66813 standard; DNA; 10035 BP.
XX
AC ABZ66813;
XX
DT 27-OCT-2003 (revised)
DT 21-MAR-2003 (first entry)
XX
DE Orthosomycin biosynthetic gene cluster SEQ ID NO 282.
XX
KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Micromonospora carbonacea; aurantiaca.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
PR 30-MAR-2001; 2001US-0279709P.
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Zazopoulos E, Staffa A;
XX
DR WPI; 2003-058435/05.
XX
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
PT cluster, by detecting presence of nucleic acid sequence corresponding to
PT 17 of flambamycins protein families.
XX
PS Example 1; Page 506-511; 51lpp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
CC ABP99362). The method is useful for identifying an orthosomycin
CC biosynthetic gene, gene fragment or gene cluster, especially an
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
CC gene fragment or gene cluster. The method is useful for detecting the
CC presence of any organism that contains DNA for the production of
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
CC orthosomycins) regardless of the level at which genes for orthosomycin
CC production are expressed by the organism or the amount of orthosomycin
CC produced by the organism. This allows for the detection of new
CC orthosomycin natural products, not produced by the organism. (Updated on
CC 27-OCT-2003 to standardise OS field)
XX
```

SQ Sequence 10035 BP; 1396 A; 3516 C; 3657 G; 1466 T; 0 U; 0 Other;  
Query Match 59.7%; Score 298.4; DB 7; Length 10035;  
Best Local Similarity 76.1%; Pred. No. 6.1e-46;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
QY 1 GGTACCCGACCTGTCCCGGAACAACGAGTCGAGATACGGGAGAGGACACCCCGGT 60  
DB 3742 GATATCCGAGCGTCTCGGTGAACAGCGAGTCAGGTACGGGACACGAAGACCCCGCG 3683  
QY 61 AGTCGGGTGAGCGTGGCGGCGAAGCGGTACCGCCCTTCAGCGGTACGCGGGCGG 120  
DB 3682 GGTGCGGAAAGACCCGCGGGGTGAAGAGTACCGCCCTCGACGCGAGTTCGGTTGCTA 3623  
QY 121 ACACC-----GGCGGGTCACTGCTCAGCTGACGCGGGGACGTACAGGATCCACTGTCC 176  
DB 3622 CTGTACGGCCCACTCAATCCCGACGTGAACGTGCGGAACGTACAGGATCCAGGCC 3563  
QY 177 GCCAGCCCGCGGGAACCTCTGCTCTCTCGCCATGATCTGCGGGGTGTTCCAGGCCAA 236  
DB 3562 GCGGCGCTGCGGGAACGCGTCTCTCTCGCCATGATCTGCGGGGTGTTCCAGCGAA 3503  
QY 237 GAGCAGCGGTAGTCCACCGGTTCGGGCGTGAACCGGTTCGGGGTGCACCGGATGTG 296  
DB 3502 GAGCAGCGGTAGTCCCGGGTTCGGTTCGGGAACCTGTCGCGCATGATCTGCGGGGTG 3443  
QY 297 CTTGCGGGGTGAGCGGCGCTGCTTGGCGGCGTGTGTCGACACCCAGGAGACAG 356  
DB 3442 CTTGCGGGGTGAGCGGCGCTGCTTGGCGGCGTGTGTCGACACCCAGGAGACAG 3383  
QY 357 GTCCGGAACGATGCGCGGAGAGTTGTCACGTTGGCGGTCTTTGCGCGTGGCGCGTACGC 416  
DB 3382 GTCCGCGCGATGCGCGGAGAGTTGTCACGTTGGCGGTCTTTGCGCGTGGCGCGTACGC 3323  
QY 417 CACCAACCCCTTGCCTCGGCGTTCGAGGAGTTACGAGGCGAGCAGGTGCGTGGCGAT 476  
DB 3322 CACCAACCCCTTGCCTCGGCGTTCGAGGAGTTACGAGGCGAGCAGGTGCGTGGCGAT 3263  
QY 477 GCGCTGAGCTCGGCGGGAAC 498  
DB 3262 CCGGCGACCTGGGAGCGAAC 3241  
RESULT 5  
ABL50562/c  
ID ABL50562 standard; DNA; 11115 BP.  
XX AC ABL50562;  
XX DT 18-JUN-2002 (first entry)  
XX DE Micromonospora carbonacea everninomicin locus nucleotide contig 8.  
XX KW Micromonospora carbonacea; antibiotic; everninomicin; biosynthesis;  
XX XW gene cluster; genetic manipulation; contig; gene; ds.  
XX OS Micromonospora carbonacea.  
XX PN WO20015180-A2.  
XX PD 02-AUG-2001.  
XX PF 29-JAN-2001; 2001WO-CA000128.  
XX PR 27-JAN-2000; 2000US-017711P.  
XX PA (ECOP-) ECOPIA BIOSCIENCES INC.  
XX PA (FARN/) FARNET C.  
XX PI Statia A, Zazopoulos E, Mercure S, Nowacki P;  
XX WPI: 2001-476185/51.  
DR P-PSDB; ABB06922, ABB06923, ABB06924, ABB06925, ABB06926, ABB06927,  
DR

DR ABB06928, ABB06929, ABB06930.  
XX Novel isolated gene cluster encoding polypeptides involved in  
PT everninomicin biosynthesis useful for construction of everninomicin  
PT overproducing strains, and to allow chemical modifications of  
PT everninomicin to enhance certain properties.  
XX Claim 1; Fig 1; 181pp; English.  
XX ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora  
CC carbonacea everninomicin biosynthetic locus gene cluster. The contigs  
CC encode the protein sequences designated ORF (open reading frame) 1 to 49,  
CC given in ABB06981 to ABB06930. The gene cluster is useful for the  
CC construction of the everninomicin antibiotic in overproducing strains,  
CC and to allow chemical modifications of everninomicin to enhance certain  
CC properties via genetic manipulation or combinatorial biosynthesis. The  
CC gene cluster can be used to produce genetic systems and genes encoding  
CC novel enzyme activities, and avoid the problems of low yield and quality  
CC of everninomicins produced by chemical synthesis  
XX Sequence 11115 BP; 1521 A; 3924 C; 4057 G; 1613 T; 0 U; 0 Other;  
SQ Query Match 59.7%; Score 298.4; DB 4; Length 11115;  
Best Local Similarity 76.1%; Pred. No. 6.1e-46;  
Matches 382; Conservative 0; Mismatches 116; Indels 4; Gaps 1;  
QY 1 GGTACCCGACCTGTCCCGGAACAACGAGTCGAGATACGGGAGAGGACACCCCGGT 60  
DB 3749 GATATCCGAGCGTCTCGGTGAACAGCGAGTTCAGGTACGGGACACGAAGACCCCGCG 3690  
QY 61 AGTCGGGTGAGCGTGGCGGCGAAGCGGTACCGCCCTTCAGCGGTGAGCGGGCGGCGG 120  
DB 3689 GGTGCGGAAAGACCCGCGGGGTGAAGAGTACCGCCCTCGACGCGGAGTTCGGTTGCTA 3630  
QY 121 ACACC-----GGCGGGTCACTGCTCAGCTGTCAGCGGGGAGCTACAGGATCCACTGTCC 176  
DB 3629 CTTGCAAGCGCCCACTCAATCCCGACGTGAACGTGCGGAACGTACAGGATCCAGCCCC 3570  
QY 177 GCCAGCCCGCGGGAACCTCTGCTCTTCGCCATGATCTCTGCGCGTGGTTCCAGCGCAA 236  
DB 3569 GCGGCGCTGCGGGAACGCGTCTCTTCGCCATGATCTCTGCGCGTGGTTCCAGCGAA 3510  
QY 237 GAGCAGCGGTAGTCCACCGGTTCGGGCGTGAACCGGTTCGGGGTGCACCGGATGTG 296  
DB 3509 GAGCAGCGGTAGTCCCGGGGTGCGTTCGGGAACCTCTCGCGCATGAACCGGGATGTG 3450  
QY 297 CTTGCGGGGTGAGCGGCGCTGCTTGGCGGCGTGTGTCGCGCGTGTGTCACACCGAGACCCAG 356  
DB 3449 CTTGCGGGGTGAGCGGCGCTGCTTGGCGGCGTGTGTCGTCACACCGAGACCCAG 3390  
QY 357 GTCCGGAACGATGCGCGGAGAGTTGTCACGTTGGCGGTCTTTGCGCGTGGCGCGTACGC 416  
DB 3389 GTCCGCGCGATGCGCGGAGAGTTGCGCACGTTGCGCGTCTTTGGCGGTGCGCCCGTAGGC 3330  
QY 417 CACCAACCCCTTGCCTCGGCGTTCAGGAGTTACAGCGGCGAGAGGTGCGTGGCGAT 476  
DB 3329 CACCAACCCCTTGCCTCGGCGTTCAGGAGTTACAGCGGCGAGGTCGTCGCGAT 3270  
QY 477 GCGCTGAGCTCGGCGGGAAC 498  
DB 3269 CCGGCGACCTGGGAGCGAAC 3248  
RESULT 6  
ABZ66784/c  
ID ABZ66784 standard; DNA; 1251 BP.  
XX AC ABZ66784;  
XX DT 27-OCT-2003 (revised)  
XX DT 21-MAR-2003 (first entry)  
XX DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 230.

XX Orthosomycin; biosynthesis; everniminomicin; avilamycin; gene; ds.  
XX Micromonospora carbonacea; aurantiaica.  
XX  
PN WO200279505-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-CA000432.  
XX  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-0279709P.  
XX 20-APR-2001; 2001US-0285214P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX WPI; 2003-058435/05.  
XX P-PSDB; ABP99321.  
XX  
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene  
XX cluster, by detecting presence of nucleic acid sequence corresponding to  
XX 17 of flamamycins protein families.  
XX  
XX Claim 16; Page 380; 51pp; English.  
XX  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everniminomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everniminomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of orthosomycin  
XX produced by the organism. This allows for the detection of new  
XX orthosomycin natural products, not produced by the organism. (Updated on  
XX 27-OCT-2003 to standardise OS field)  
XX  
XX Sequence 1251 BP; 188 A; 431 C; 439 G; 193 T; 0 U; 0 Other;  
XX  
XX Query Match 51.3%; Score 256.6; DB 7; Length 1251;  
XX Best Local Similarity 81.2%; Pred. No. 3.1e-38;  
XX Matches 298; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
XX  
QY 132 TCAGCTGTCACGTGACGGGGACGTACAGATCCACTGTCCGCCAGCCCGCGGAA 191  
DB 1251 TCAATCCCGACGTGACGTGCGGAACGTACAGATCCAGGCCCGCCGCGCTGCCGGA 1192  
QY 192 CTCTGCTCTTGGCCATGATCTCGTGGGGTGGTCCAGGCGAAGACGCGTATGTC 251  
DB 1191 CGCTGTCTCTTGGCCATGATCTCGTGGGGTGGTCCAGGCGAAGACGCGTATGTC 1132  
QY 252 CACCGCTGCGGGGTGAACCGCTGCGGGGTGCGACCGGATGTGCTGCGGGGTGAG 311  
DB 1131 CGGCGGGTGTGTCGGGAACCTGTCGCGCGATGAACCGGATGTGCTGCGGGGTGAG 1072  
QY 312 CCGGCCCTGTGTCGCGGGTGTGTCGACACCCAGGACACAGTCCGAGCCGATGCC 371  
DB 1071 GCGGCCCTGTGTCGCGGGTGTGTCGACACCCAGGACACAGTCCGAGCCGATGCC 1012  
QY 372 GCAGAAATGTCACGTGCGGTCTTTCGCGGTGCGGCGCTACGCCACACCCGCTGTC 431  
DB 1011 GCAGAAATGTCACGTGCGGTCTTTCGCGGTGCGGCGCTACGCCACACCCGCTTTC 952  
QY 432 CTGCGGCTTGAGCGAGTTCAGGCGGACGAGTGTGCGGTGCGGATGCCCTTCGACGTCGC 491  
DB 951 CTCGCCCTGTAACGTTTCAGGACAAACGCCCGCAGGTGTCGCGGATCCGCGGACCTGGGC 892

QY 492 GCGGAAC 498  
DB 891 AGCGAAC 885  
RESULT 7  
AAV34002/c  
ID AAV34002 standard; DNA; 1401 BP.  
XX  
XX AAV34002;  
XX  
XX 27-AUG-2003 (revised)  
XX 25-MAR-2003 (revised)  
XX 25-JAN-1999 (first entry)  
XX  
XX S. peucetius dnrX gene.  
XX  
XX dnrU; dnrX; recombinant; microorganism; doxorubicin; daunorubicin;  
XX metabolism; baumycin; secretion; anti-tumour compound; ds.  
XX Streptococcus.  
XX  
XX Key Location/Qualifiers  
XX CDS 190..1401  
XX FT /\*tag= a  
XX FT /product= "dnrX gene product"  
XX  
XX WO9839458-A1.  
XX  
XX 11-SEP-1998.  
XX  
XX 05-MAR-1998; 98WO-US003938.  
XX  
XX 06-MAR-1997; 97US-00812412.  
XX 28-JUL-1997; 97US-00901306.  
XX  
XX (PAAA ) PHARMACIA & UPJOHN SPA.  
XX  
XX Filippini S, Lomovskaya N, Fonstein L, Colombo AL, Hutchinson RC;  
XX Otten S;  
XX WPI; 1998-506367/43.  
XX P-PSDB; AAW68561.  
XX  
XX Production of doxorubicin and daunorubicin - using a recombinant  
XX daunorubicin producing microorganism in which at least one dnrU or dnrX  
XX metabolism gene has been inactivated.  
XX  
XX Disclosure; Page 29-32; 48pp; English.  
XX  
XX This sequence represents the dnrX gene from Streptomyces peucetius. The  
XX invention relates to a recombinant microorganism which produces  
XX doxorubicin (DoR), in which at least 1 daunorubicin (DaR) metabolism gene  
XX selected from dnrU and dnrX, has been inactivated. The microorganism is  
XX preferably S. peucetius WH1658, WH1654 and WH1662. The function of the  
XX DaR gene is inactivated by insertion of a gene (preferably  
XX neomycin/kanamycin resistance gene aphII) into the daunorubicin  
XX metabolism gene. The blocking of the function of at least one gene of DaR  
XX metabolism can increase DaR and DoR production levels and cause the  
XX disappearance of baumycin-like products resulting in DaR and DoR  
XX secretion directly into the culture medium. The products can be used in  
XX the production of anti-tumour compounds. (Updated on 25-MAR-2003 to  
XX correct PI field.) (Updated on 27-AUG-2003 to correct OS field.)  
XX  
XX Sequence 1401 BP; 216 A; 491 C; 473 G; 221 T; 0 U; 0 Other;  
XX  
XX Query Match 30.0%; Score 149.8; DB 2; Length 1401;  
XX Best Local Similarity 65.3%; Pred. No. 9.6e-19;  
XX Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
XX  
QY 164 GATCCATGTCGCGACCCGCGGACCTCTCTCTTCGCGCATGATCTGTCGCGGT 223  
DB 1390 GATCCAGCGGCCCTCTCTCTTCGCGCATGATCTGTCGCGGTCTGCTAGGT 1331

QY 224 GGTTCAGGCGAAGAGAGAGCGCGTAGTCCACCGCGTCCGGCGGTGAACGCGTCCGGGGTGC 283  
 Db 1330 GGTTCAGGCGAAGAGAGAGCGCATAGTCCGGGTAGGGGGCCCGGAACCTCCTCCGGCGAGC 1271  
 QY 284 GCACCGGGATGTGGTGC CGGGGGTGAAGCGCGCCCTCTTTGGCGGGTCTGTCGCGACA 343  
 Db 1270 GGATGGGGATGTGGAGACCGGGGAGCGACGGCGGATCTTGGCGGGTGTGTCTGTAGA 1211  
 QY 344 CCCAGGAGACAGATCCCGGACCGATGCCGAGAAAGTTTCGTACGGTGGCGCTTTTCGCGG 403  
 Db 1210 CGCAGCGCAGCAGATCGGGGGCCGATCCCGCATAGTTGAGCACTGTGGCGCTCTTGCGGG 1151  
 QY 404 TCGCGCGTACGCCACACCGCTTGGCTCGGCTTTAGCGAGTTTCAGCAGGCGCGAGCA 463  
 Db 1150 TCGCGCGTACCGACGACGATAGAGCGGAGTCCGCGAGTCCGCGAGCGGCGACCA 1091  
 QY 464 GGTTCGGTCCGATCCCTTCGACGTCGCGCGGCGAACCT 500  
 Db 1090 GGTTCACAGCAGACCGCTTGACCTGGCGCGCGAACCT 1054

## RESULT 8

AAA39283/c  
 ID AAA39283 standard; DNA; 16020 BP.

XX AC AAA39283;

XX DT 08-SEP-2000 (first entry)

XX DE Streptomyces nogalater nogalamycin biosynthesis gene cluster SEQ ID NO:1.

XX KW Streptomyces nogalater; nogalamycin biosynthesis; antibiotic;

XX KW anthracycline biosynthetic pathway; gene cluster; drug screening;

XX KW antibiotic; antitumour antibiotic; anthracycline; ds.

XX OS Streptomyces nogalater.

XX FN WO200024775-A1.

XX PD 04-MAY-2000.

XX PF 20-OCT-1999; 99WO-FI000870.

XX PR 23-OCT-1999; 99FI-00002295.

XX PA (GALI-) GALILAEUS OY.

XX PI Ylihanko K, Torkkell S, Palmu K, Hakala J;

XX DR WPI; 2000-350696/30.

XX DR P-ESDB; AAY91055, AAY91056, AAY91057, AAY91058, AAY91059, AAY91060,

XX DR AAY91061, AAY91062, AAY91063, AAY91064, AAY91065, AAY91066, AAY91067,

XX DR AAY91068, AAY91069, AAY91070, AAY91071.

XX PT Isolated and purified DNA fragment for obtaining novel hybrid antibiotics  
 PT comprises the gene cluster for the anthracycline biosynthetic pathway of  
 PT the bacterium Streptomyces nogalater.

XX FS Claim 2; Page 32-40; 59pp; English.

XX CC The present sequence represents the nogalamycin biosynthesis gene cluster  
 CC isolated from Streptomyces nogalater. Nogalamycin is an anthracycline  
 CC antibiotic, so the nogalamycin biosynthetic pathway is also known as the  
 CC anthracycline biosynthetic pathway. DNA fragments, plasmids and process  
 CC from the present invention are useful for obtaining novel hybrid  
 CC antibiotics, such as anthracyclines (antitumour antibiotics) or  
 CC anthracyclonones

XX SQ Sequence 16020 BP; 2295 A; 6011 C; 5425 G; 2189 T; 0 U; 100 Other;

Query Match 27.0%; Score 135; DB 3; Length 16020;

Best Local Similarity 61.5%; Pred. No. 4e-16;

Matches 216; Conservative 0; Mismatches 135; Indels 0; Gaps 0;  
 QY 150 GCGGGGACGTACAGATCCACTGTCGCCAGCCCGCGGAACTCCTGCTCTTCCGCAT 209  
 Db 6323 GCGGGGATGGGACGATGAACCGGCCCGCTTGTCCCGGAAGCCGCTCTCTTCTCGAG 6264  
 QY 210 GATCTCGTCGGCGTGTTCAGGCGAAGAGCAGCGGTAGTCCACCGGTGCGGGGTGAA 269  
 Db 6263 GATCTCCGTGGGTAGTTCAGGCGGAGCAACAGTAGTAGTCGGGGATGTGTTCTTGGC 6204  
 QY 270 CCGCTCCGGGTGCGCACCGGAGTGTGCTGCGGGGTGAGCCGGCCCTGCTTGGCCGG 329  
 Db 6203 GTGCTCGGAGCGTCAACCGGTATGTGGGTGCGGCGCAGCACCTTGCCTTCACTC 6144  
 QY 330 CCGCTGTCGACACCCAGGAGACAGTCCGAGCCGATGCCGAGAAAGTTGTCACCGT 389  
 Db 6143 GGTGTGTCTGCTGAGTATTCAGTCTCTTCCAGCCGAGCCCGACACCGTGAGATGTT 6084  
 QY 390 GCGCTCTTCCGCGTCCGCGGTAGCCACACCGCTTGGCCCTCGGCTTGAAGCGAGTT 449  
 Db 6083 GTTGCCCTTGGCGGAGCACCGTAGCCGACGATGCGTTGCCCTGGGCCACGAGGAGCG 6024  
 QY 450 CAGCAGGCGCAGCAGTCCGTGCGGATGCCCTCGACGTGCGGGCGAACCT 500  
 Db 6023 CAGCAGTTCGCGCAGTCCGTGCGGACCGCGCTCGATCCGCTGCGCGAACCT 5973

## RESULT 9

AAS08699  
 ID AAS08699 standard; DNA; 12152 BP.

XX AC AAS08699;

XX DT 11-SEP-2003 (revised)

XX DT 26-SEP-2001 (first entry)

XX DE M. carbonacea DNA encoding Everninomicin biosynthetic enzymes.

XX KW Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation;  
 KW ds.

XX OS Micromonospora sp. ATCC 39149.

XX PH Key Location/Qualifiers  
 CDS complement(189..1064)

FT /tag= a

FT /product= "ORF1"

FT complement(1069..1073)

FT /tag= b

FT 1184..2767

FT /tag= c

FT /product= "ORF2"

FT complement(2683..3753)

FT /tag= d

FT /product= "ORF3"

FT complement(3766..4276)

FT /tag= e

FT /product= "ORF4"

FT complement(4280..4284)

FT /tag= f

FT complement(4526..5368)

FT /tag= g

FT /product= "ORF5"

FT complement(5392..6147)

FT /tag= h

FT /product= "ORF6"

FT complement(6152..6156)

FT /tag= i

FT complement(6194..7282)

FT /tag= j

FT /product= "ORF7"

FT complement(7280..8133)

FT /tag= k

FT	/product= "ORF8"
RBS	complement(8141. .8145)
FT	/*tag= l
FT	complement(8254. .9318)
CDS	/*tag= m
FT	/product= "ORF9"
FT	complement(9324. .9328)
FT	/*tag= n
PT	9568. .9571
FT	/*tag= p
FT	9575. .10504
CDS	/*tag= o
FT	/product= "ORF10"
FT	10584. .11585
CDS	/*tag= q
FT	/product= "ORF11"
FT	

WO200151639-A2.

19-JUL-2001.

12-TAN-2001: 2001WO-US001187.

12-JAN-2000: 2000US-0175751P.

(SCHE ) SCHERTING CORP.

Hosted TJ. Horan AC. Wang TX:

WBT: 2001-442147/A7

P=PSDB; AAU04901, AAU04902, AAU04903, AAU04904, AAU04905, AAU04906, AAU04907 AAU04908 AAU04909 AAU04910 AAU04911

New nucleic acid molecules encoding everninomicin pathway gene products, useful for improving yields of everninomicin, to produce new everninomicin and as probes to identify homologous sequences.

Example 1: Fig 12: 109pp: English.

The sequence encodes 11 proteins comprising enzymes of the everniminomicin antibiotic biosynthetic pathway. A vector comprising a *M. carbonacea* everniminomicin biosynthetic pathway resistance gene product is useful for selecting for a transfected or transformed host cell. An integrative version of the vector is useful for introducing a everniminomicin pathway gene (a bottle-neck gene) into an actinomycete of the genus *Micromonospora*. The DNA encoding the biosynthetic proteins is useful for synthesising novel everniminomicin-related compounds, arising from modifications of the DNA sequence designed to change glycosyl and modified orellanic acid groups contained in everniminomicin, for expressing functional or mutant everniminomicin biosynthetic enzyme for evaluation, diagnosis and preferably biosynthesis of everniminomicin or other secondary metabolic products, improving the yield of everniminomicins and to produce novel everniminomicins and also as a hybridisation probe to identify homologous sequences. The encoded polypeptides are useful for combinatorial biosynthesis to generate libraries of orthomycins, e.g. everniminomicin analogues/homologues and drug discovery. The DNA encoding the integrase allows for increasing a given gene dosage. The integrative vector can be used to permanently integrate copies of a heterologous gene of choice into chromosomes of different hosts and to integrate genes which increase the yield of known products or to generate novel products such as hybrid antibiotics or other novel secondary metabolites. The vector can also be used to integrate antibiotic resistance genes in order to carry out bioconversions with compounds to which the strain is normally sensitive and is thus useful in fermentation processes involving e.g. Streptomyces antibioticus. (Updated on 11-SEP-2003 to standardise OS field).

Sequence 12152 BP: 1675 A: 4502 C: 4475 G: 1500 T: 0 U: 0 Other:

every March 26.2%: score 131: DB 5: Length 12152:

```

Best Local Similarity 100.0%; Pred. NO. 2.2e-13;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy	1	G	T	A	C	C	G	A	C	G	T	G	T	C	C	G	A	C	A	C	A	C	A	C	A	C	A	C	C	C	C	G	T	60	
Db	12022	G	T	A	C	C	G	A	C	G	T	G	T	C	C	G	A	C	A	C	A	C	A	C	A	C	A	C	C	C	C	G	T	12081	
Qy	61	A	T	C	C	G	G	T	A	G	A	C	G	T	G	G	C	G	A	A	G	G	C	T	T	C	A	C	G	T	A	G	C	G	120
Db	12082	A	T	C	C	G	G	T	A	G	A	C	G	T	G	G	C	G	A	A	G	G	C	T	T	C	A	C	G	T	A	G	C	G	12141
Qy	121	A	C	A	C	C	G	C	G	G	131																								
Db	12142	A	C	A	C	C	G	C	G	G	12152																								

RESULT 10  
ABZ66779/C

ID ABZ66779 standard; DNA; 603 BP.

AC ABZ66779;

27-OCT-2003 (revised)

XX  
XX  
CO07-PM-12 ID  
JSTJ 08111

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

[illegible]US  
MICROMESOPOTA CARPOTACEA, ATTRICATA  
XX[illegible]

XX  
XX  
PL  
:2002-10-01

28-MAR-2002; 2002WU-CA00043Z;  
PF  
XX

PR 28-MAR-2001; 2001US-0275095F.  
PR 30-MAR-2001; 2001US-0279709P.

PR 20-APR-2001; 200105-0285214F.  
XX

PA (ECUP-) ECOPLA BIOSCIENCES INC.  
YY

PI Farnet CM, Lazopoulos E, Stalla A, VV

DR WPI; 2003-058435/05.  
 DR B-PQNB: ABP9931C

2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2808 2809 2810 2811 2812 2813 2814 2815 2816 2817 2

cluster, by detecting presence of nucleic acid sequence corresponding to

XX  
50 01-24-16. D-220 258-270. 511mm. Eng[?]ish

xx The invention relates to identifying orthosomycin biosynthetic genes and  
cc its fragment/gene cluster (ABZ56670-ABZ656813), comprising detecting the  
cc presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
cc ABP99362). The method is useful for identifying an orthosomycin  
cc biosynthetic gene, gene fragment or gene cluster, especially an  
cc evernimicin-type or avilamycin-type orthosomycin biosynthetic gene,  
cc gene fragment or gene cluster. The method is useful for detecting the  
cc presence of any organism that contains DNA for the production of  
cc orthosomycins (both evernimicin-type orthosomycins and avilamycin-type  
cc orthosomycins) regardless of the level at which genes for orthosomycin  
cc production are expressed by the organism or the amount of orthosomycin  
cc produced by the organism. This allows for the detection of new  
cc orthosomycin natural products, not produced by the organism. (Updated on  
cc 27-Oct-2003 to standardise OS field)

XX  
SQ  
sequence 603 BP: 77 A: 224 C: 197 G: 105 T: 0 U: 0 Other:

Query Match 24.8%: Score 124: DB 7: Length 603;

Best Local Similarity	100.0%;	Freq: NO: 3.5e-14;
Matches	124;	Conservative
0:	Mismatches	0:
Indels	0:	Gaps
0:		

1 GGTACCCCGACCGTGTCCTCGGACACACGAGTCGAGATACGGCGAGAGGAACACCCCCGGT 60

Db 124 GGTACCCGACCGTGTCCCGGACACAGTGTGAGATACGGGAGAGGAACACCCCGGT 65
Qy 61 AGTCGGGTAGACGGTGGCGCGAAGCGGTACGGCCCTTCACCGGTAGCGGGCGCGG 120
Db 64 AGTCGGGTAGACGGTGGCGCGAAGCGGTACGGCCCTTCACCGGTAGCGGGCGCGG 5
Qy 121 ACAC 124
Db 4 ACAC 1
RESULT 11
ABZ66809/c
ID ABZ66809 standard; DNA; 15240 BP.
XX
AC ABZ66809;
XX
DT 27-OCT-2003 (revised)
DT 21-MAR-2003 (first entry)
XX
Orthosomycin biosynthetic gene cluster SEQ ID NO 278.
XX
Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Micromonospora carbonacea; africana.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
PR 30-MAR-2001; 2001US-0279709P.
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Zazopoulos E, Staffa A;
XX
DR WPI; 2003-058435/05.
XX
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
PT cluster, by detecting presence of nucleic acid sequence corresponding to
PT 17 of flambamycins protein families.
XX
XX Example 10; Page 450-458; 51pp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
CC ABP99362). The method is useful for identifying an orthosomycin
CC biosynthetic gene, gene fragment or gene cluster, especially an
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic
CC gene fragment or gene cluster. The method is useful for detecting the
CC presence of any organism that contains DNA for the production of
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
CC orthosomycins) regardless of the level at which genes for orthosomycin
CC production are expressed by the organism or the amount of orthosomycin
CC produced by the organism. This allows for the detection of new
CC orthosomycin natural products, not produced by the organism. (Updated on
CC 27-OCT-2003 to standardise OS field)
XX
SQ Sequence 15240 BP; 2131 A; 5448 C; 5454 G; 2207 T; 0 U; 0 Other;
Query Match 23.6%; Score 118.2; DB 7; Length 15240;
Best Local Similarity 53.3%; Pred. No. 4.6e-13;
Matches 249; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
Qy 33 AGATACCGGAGAGGACACCCCGGTAGTCGGGTAGACGGTGGCGCGGAGGCGGTAC 92
Db 12684 AGAGATCGATCCAGGGCAGCAGCATGTCTCCGGGACCGGGCTTCCGGGTCAGTATGA 12625

Qy 93 GCGCCTTCACGGTCAGCGGGCGGGCGGACACCGCGGGTCAGTCGTCACTGTACGCG 152
Db 12624 AAGTTGATCACCTCGGTGTGGGGGGGTGTTCATCCGGTCGGCGACCGTCTCGACCG 12565
Qy 153 GGGGAGGTACAGATCACTGTCCCGCAGCCCGCGGGAACCTCTGCTCTTTCGCCATGAT 212
Db 12564 GGGCAGCGGAAGATCAACCCGGTGCCTGTGCCAGCATCTCCGCTCCCGGGCGACGAT 12505
Qy 213 CTCGTGGCGGTGTTCACGGCGAAGAGCAGCGGTAGTCCACCGGTCCGGGGGTGAACGC 272
Db 12504 CTCGTCCCGAAGTGCAGGGCAGGACAGGTAGTAGTCCGGCGGGCCGCCGCGACTC 12445
Qy 273 GTCGGGGTGGCACCGGATGTGTCGGGGGTGAGCCGCCCTTGTTCGCCCGGCGT 332
Db 12444 CTGCTCGCTGATGATTTTCGATGTCCGTGCCGAGGGTGCCTGCCGCCACCTTGTCCGGATT 12385
Qy 333 CGTGTGCGACACCCAGGAGACAGGTCCGACCGATGCCAGAGTTCGTCACTGTGGC 392
Db 12384 GCGCTCCGGGGTACCGGATGAGCTCGCGTCTATGCCGCGAGAACTGCAGGAGGTGT 12325
Qy 393 GCTCTTTCGGCGTCGCGCCGTACGCCACCCCGCTTGCCTCGGCTTTCGCGGAGTTT 452
Db 12324 GCCCTTCGTGCGCGCGCGGTAGACGTGCACCGTGCCTCGCGCGCTTGCCTCGCGCAG 12265
Qy 453 CAGGGCGAGCAGGTGCGGTGCGGATGCCCTCGAGCTCGGGCGGAACC 499
Db 12264 CAGGGCGCTCACCTGTACGGTGTGGCGACCTCTCGCGAAGC 12218
RESULT 12
ABZ66730/c
ID ABZ66730 standard; DNA; 1281 BP.
XX
AC ABZ66730;
XX
DT 27-OCT-2003 (revised)
DT 21-MAR-2003 (first entry)
XX
Orthosomycin biosynthetic polynucleotide SEQ ID NO 122.
XX
Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.
XX
OS Micromonospora carbonacea; africana.
XX
PN WO200279505-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-CA000432.
XX
PR 28-MAR-2001; 2001US-0279095P.
PR 30-MAR-2001; 2001US-0279709P.
PR 20-APR-2001; 2001US-0285214P.
XX
PA (ECOP-) ECOPIA BIOSCIENCES INC.
XX
PI Farnet CM, Zazopoulos E, Staffa A;
XX
DR WPI; 2003-058435/05.
XX
DR P-PSDB; ABP99267.
XX
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
PT cluster, by detecting presence of nucleic acid sequence corresponding to
PT 17 of flambamycins protein families.
XX
XX Example 10; Page 275-276; 51pp; English.
XX
CC The invention relates to identifying orthosomycin biosynthetic genes and
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
CC ABP99362). The method is useful for identifying an orthosomycin
CC biosynthetic gene, gene fragment or gene cluster, especially an

```
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,
CC gene fragment or gene cluster. The method is useful for detecting the
CC presence of any organism that contains DNA for the production of
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type
CC orthosomycins) regardless of the level at which genes for orthosomycin
CC production are expressed by the organism or the amount of orthosomycin
CC produced by the organism. This allows for the detection of new
CC orthosomycin natural products, not produced by the organism. (Updated on
CC 27-OCT-2003 to standardise OS field)
XX
SQ Sequence 1281 BP; 182 A; 440 C; 469 G; 190 T; 0 U; 0 Other;

Query Match      22.5%; Score 112.8; DB 7; Length 1281;
Best Local Similarity 56.5%; Pred. No. 5.5e-12;
Matches 210; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 128 CCGGTGAGTCTGTCAGTGTACGCGGGGACGTACAGGATCCACTGTCGCGACGCCCGC 187
Db 1276 CCGGTGCGCGACCGCTCGACCCCGGGCAGCGGAAGATCAACCGGTGCGCGTGGCCA 1217

QY 188 GGAATCTCTGCTTCGCGCATGATCTCTGCGCGTGTTCAGGCGAAGACGCGGT 247
Db 1216 GCATCTCCGCTCCCGGGGACGATCTCTCCGGAAGTGCCAGGCGACGAGTGT 1157

QY 248 AGTCCACCGCGTCCGCGGTGAACCGCTCCGGGTGCGCACCGGGATGTCGTCGCGGGG 307
Db 1156 AGTCCCGGGCGGCGCGCGACTCTGCTCGCTGATGATTTTCGATGTCGTCGCGAGG 1097

QY 308 TGAGCGCGCCCTGCTTGCGCGCGGTCTGTCTGCAACCCAGGAGACCGGTCCGACCGA 367
Db 1096 TCGCGCGCGCCACTTGTCCGGATTCCGCTCCGCGCGTACCGGATGAGCTCGCGTCTA 1037

QY 368 TCGCGCAGAGTTGTCAGCGTGGCGCTCTTGCGCGTCCGCGCGTACGCCACACCCGCT 427
Db 1036 TCGCGCAGAACTCAGAGAGGTGTTCCTTCTGTCAGCGCGGTAGAGTGCACCGTGC 977

QY 428 TGCCCTCGCGCTTGAGCGATTTCAGAGGCGAGCGTTCGTCGCGATGCCCTCGACGT 487
Db 976 GGCCTCGCGCCCGCAGTCTCGCGCAGCGGCGTCACTCGTACCGTCTCGCGCACCT 917

QY 488 CGCGCGGACACC 499
Db 916 GCTCGCGGAGC 905

RESULT 13
ID AAA63349 standard; DNA; 42000 BP.
XX
AC AAA63349;
XX
XX 06-MAR-2001 (first entry)
XX Streptomyces globisporus C-1027 gene cluster ORF -7-25.
XX Enediyne C-1027 biosynthesis gene cluster; apoprotein; chromophore;
XX cancer; ds.
XX Streptomyces globisporus.
XX Key Location/Qualifiers
FH complement(8. .658)
FT /*tag= a "ORF -7 protein"
FT complement(930. .1478)
FT /*tag= b
FT /product= "ORF -6 protein"
FT complement(1649. .2713)
FT /*tag= c
FT /product= "ORF -5 protein"
FT complement(2850. .3237)
FT /*tag= d
FT /product= "ORF -4 protein"
FT
```

```
FT complement(3442. .4971)
FT /*tag= e
FT /product= "ORF -3 protein"
FT 5982. .7479
FT /*tag= f
FT /product= "glycerol phosphate transporter"
FT complement(7573. .9900)
FT /*tag= g
FT /product= "ABC transport/UvrA-like protein"
FT complement(9982. .11349)
FT /*tag= h
FT /product= "Na+/H+ transporter"
FT complement(11351. .12835)
FT /*tag= i
FT /product= "hydroxylase/halogenase"
FT 13012. .14079
FT /*tag= j
FT /product= "dNDP-glucose synthase"
FT complement(14212. .14643)
FT /*tag= k
FT /product= "CagA"
FT complement(14690. .15922)
FT /*tag= l
FT /product= "aminotransferase"
FT complement(15919. .16653)
FT /*tag= m
FT /product= "N-methyl transferase"
FT complement(16653. .17924)
FT /*tag= n
FT /product= "C-methyl transferase"
FT complement(18031. .19191)
FT /*tag= o
FT /product= "epoxide hydrazase"
FT complement(19267. .19929)
FT /*tag= p
FT /product= "anthranilate synthase II"
FT complement(19926. .21407)
FT /*tag= q
FT /product= "anthranilate synthase I"
FT complement(21424. .22878)
FT /*tag= r
FT /product= "coenzyme F390 synthetase"
FT complement(22875. .23546)
FT /*tag= s
FT /product= "iron-sulphur flavoprotein"
FT complement(23566. .24702)
FT /*tag= t
FT /product= "O-acyl transferase"
FT complement(24986. .25564)
FT /*tag= u
FT /product= "epimerase"
FT 25815. .27170
FT /*tag= v
FT /product= "monooxygenase"
FT 27214. .28593
FT /*tag= w
FT /product= "glycosyl transferase"
FT 28590. .29588
FT /*tag= x
FT /product= "dNDP glucose dehydratase"
FT 29632. .31197
FT /*tag= y
FT /product= "SgsB transmembrane efflux protein"
FT 31280. .32590
FT /*tag= z
FT /product= "coenzyme F390 synthetase"
FT 32809. .34392
FT /*tag= aa
FT /product= "hydroxylase"
FT complement(34458. .35294)
FT /*tag= ab
FT /product= "ORF 22 protein"
FT /transl_except= (pos:35223. .35225,aa:11e)
```





FT	/product= "epoxide hydrase"
FT	complement(19257. .19929)
FT	/*tag= p
FT	/product= "anthranilate synthase II"
FT	complement(19926. .21407)
FT	/*tag= q
FT	/product= "anthranilate synthase I"
FT	complement(21424. .22878)
FT	/*tag= r
FT	/product= "coenzyme F390 synthetase"
FT	complement(22875. .23546)
FT	/*tag= s
FT	/product= "iron-sulphur flavoprotein"
FT	complement(23566. .24702)
FT	/*tag= t
FT	/product= "O-acyl transferase"
FT	complement(24956. .25564)
FT	/*tag= u
FT	/product= "epimerase"
FT	25815. .27170
FT	/*tag= v
FT	/product= "mnoooxygenase"
FT	27214. .28593
FT	/*tag= w
FT	/product= "glycosyl transferase"
FT	28590. .29388
FT	/*tag= x
FT	/product= "dNDP glucose dehydratase"
FT	29632. .31197
FT	/*tag= y
FT	/product= "SgCB transmembrane efflux protein"
FT	31280. .32590
FT	/*tag= z
FT	/product= "coenzyme F390 synthetase"
FT	32809. .34392
FT	/*tag= aa
FT	/product= "hydroxylase"
FT	complement(34458. .35294)
FT	/*tag= ab
FT	/product= "ORF 22 protein"
FT	/transl_except= {pos:35223. .35225,aa:Ile}
FT	/transl_except= {pos:35226. .35228,aa:Ala}
FT	complement(35518. .35938)
FT	/*tag= ac
FT	/product= "ORF 18 protein"
FT	36165. .37490
FT	/*tag= ad
FT	/product= "coenzyme F390 synthetase"
FT	37559. .38938
FT	/*tag= ae
FT	/product= "type II NRPS condensation enzyme"
FT	39983. .39264
FT	/*tag= af
FT	/product= "type II peptidyl carrier protein"
FT	complement(39367. .40986)
FT	/*tag= ag
FT	/product= "aminomutase"
FT	complement(41052. .42611)
FT	/*tag= ah
FT	/product= "type II NRPS adenylation enzyme"
FT	43945. .46024
FT	/*tag= ai
FT	/product= "transmembrane transport protein"
FT	46167. .47171
FT	/*tag= aj
FT	/product= "O-methyl transferase"
FT	47228. .48484
FT	/*tag= ak
FT	/product= "F450 hydroxylase"
FT	complement(48607. .49714)
FT	/*tag= al
FT	/product= "oxidoreductase"
FT	50350. .51390

FT	/*tag= am
FT	/product= "ORF 31 protein"
FT	complement(51421. .52341)
FT	/*tag= an
FT	/product= "oxidoreductase"
FT	53241. .54074
FT	/*tag= ao
FT	/product= "ORF 33 protein"
FT	54231. .55379
FT	/*tag= ap
FT	/product= "ORF 34 protein"
FT	56026. .56880
FT	/*tag= aq
FT	/product= "proline oxidase"
FT	complement(56927. .57736)
FT	/*tag= ar
FT	/product= "ORF 36 protein"
FT	complement(57833. .58304)
FT	/*tag= as
FT	/product= "ORF 37 protein"
FT	complement(58440. .60095)
FT	/*tag= at
FT	/product= "P450 hydroxylase"
FT	complement(60092. .60621)
FT	/*tag= au
FT	/product= "ORF 39 protein"
FT	60940. .62019
FT	/*tag= av
FT	/product= "ORF 40 protein"
FT	62045. .62899
FT	/*tag= aw
FT	/product= "ORF 41 protein"
FT	complement(62787. .63164)
FT	/*tag= ax
FT	/product= "ORF 42 protein"
XX	
PN	WO200040596-A1.
XX	
XX	13-JUL-2000.
PD	
XX	06-JAN-2000; 2000WO-US000446.
PF	
XX	
PR	06-JAN-1999; 99US-0115434P.
XX	
PR	05-JAN-2000; 2000US-00477962.
XX	
PA	(REGC ) UNIV CALIFORNIA.
XX	
PI	Shen B, Liu W, Christenson SD, Standage S
XX	
DR	WPI; 2000-465947/40.
DR	P-PDB; AAB13554, AAB13555, AAB13556, AAB13557, AAB13558, AAB13559, AAB13560, AAB13561, AAB13562, AAB13563, AAB13564, AAB13565, AAB13566, AAB13567, AAB13568, AAB13569, AAB13570, AAB13571, AAB13572, AAB13573, AAB13574, AAB13575, AAB13576, AAB13577, AAB13578, AAB13579, AAB13580, AAB13581, AAB13582, AAB13583, AAB13584, AAB13585, AAB13586, AAB13587, AAB13588, AAB13589, AAB13590, AAB13591, AAB13592, AAB13593, AAB13594, AAB13595, AAB13596, AAB13597, AAB13598, AAB13599, AAB13600, AAB13601, AAB13602, AAB13603, AAB13604, AAB13605, AAB13606, AAB13607.
XX	
PT	Isolated nucleic acid comprising a nucleic acid
PT	open reading frames (ORFs) -7 to 42, excluding
PT	the production of enediyne C-1027 antitumor
PS	Claim 1; Page 78-157; 160pp; English.

```
Query Match      19.4%; Score 97; DB 3; Length 63164;
Best Local Similarity 55.1%; Pred. No. 3e-09;
Matches 190; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```



```
FT CDS 22780..24042
FT /*tag= u
FT /gene= "novU"
FT /product= "NovU"
FT /note= "encodes ABR42576, C-methyltransferase"
FT 24104..24937
FT /*tag= v
FT /gene= "novV"
FT /product= "NovV"
FT /note= "encodes ABR42577, dTDP-1-glucose synthase"
FT 24994..25617
FT /*tag= w
FT /gene= "novW"
FT /product= "NovW"
FT /note= "encodes ABR42578, dTDP-4-keto-6-deoxyglucose 3,5-epimerase"
FT XX
FT PN W02003014352-A2.
FT XX
FT XX 20-FEB-2003.
FT XX
FT XX 06-AUG-2002; 2002WO-EP008777.
FT PF
FT XX 08-AUG-2001; 2001US-0310808P.
FT PR
FT XX (UYTU-) UNIV TUEBINGEN.
FT XX
FT XX Heide L, Li S;
FT PI
FT XX WPI; 2003-268202/26.
FT DR
FT DR P-PSDB; ABR42556, ABR42557, ABR42558, ABR42559, ABR42560, ABR42561,
FT DR ABR42562, ABR42563, ABR42564, ABR42565, ABR42566, ABR42567, ABR42568,
FT DR ABR42569, ABR42570, ABR42571, ABR42572, ABR42573, ABR42574, ABR42575,
FT DR ABR42576, ABR42577, ABR42578.
FT XX
FT PT Novel polypeptide encoded by open reading frame of aminocoumarin
FT PT biosynthetic gene cluster, useful for synthesizing aminocoumarins useful
FT PT for treating bacterial infections and malignancies.
FT XX
FT PS Claim 2; Fig 4; 124pp; English.
FT XX
FT CC The present sequence is the complete sequence of the Streptomyces
FT CC spheroides novobioicin biosynthetic gene cluster. The sequences and
FT CC arrangement of the open reading frames of this gene cluster show
FT CC similarity to the biosynthetic gene clusters of the aminocoumarins
FT CC coumermycin A1, chlorobioicin and simocyclinone (see ACC58250-53). The
FT CC invention provides aminocoumarin biosynthetic nucleic acids and
FT CC polypeptides, a method for producing hybrid antibiotics using a
FT CC combination of various regions of biosynthetic gene clusters, and a
FT CC pharmaceutical composition containing an aminocoumarin compound for use
FT CC in the treatment of Gram-positive bacterial infections and of malignant
FT CC diseases. (Updated on 27-Oct-2003 to standardise OS field)
FT XX
FT SQ Sequence 25617 BP; 4135 A; 8974 C; 8665 G; 3843 T; 0 U; 0 Other;
FT
FT Query Match 19.0%; Score 95; DB 8; Length 25617;
FT Best Local Similarity 54.0%; Pred. No. 7.6e-09;
FT Matches 194; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
FT
FT QY 140 TCACGTGTACCGGGGACGTACAGGATCCACTGTCCGCGAGCCCGCGGACCTCTGCT 199
FT DB 24019 TCACCGCGGTTCGGGTATCGGAACGATGAATCGCGCGCGTCTTCCAGGAATTCCTTCT 23960
FT
FT QY 200 CCTTCGCGCATGATCTCGTCGCGTGTTCAGGCGAAGACGCGGTAGTCCACCGGT 259
FT DB 23959 CGTTGTGATGATCTCGGTGCGGTAGTTCCAGGCCAACACAGGTAGTATCGGTACAT 23900
FT
FT QY 260 CGGCGGTGAACCGGTTCGGGGTGCACACCGGGATGTGCGTGC CGGGGGT GAGCGGCCCT 319
FT DB 23899 GGC CGGCGCGCTGTTCGGGTGACCATCGGAATCTTGACCCCGCGAAGTATTTCCCT 23840
FT
FT QY 320 GCTTGGCGGGCGTTCGTCGCGACACCCAGAGACCGGTCCGAGCGATGCCGAGAAT 379
FT
```

```
Db 23839 GCTTGAGAACGGTTGTGTCTCGTACGAAACGGGATCTCCTGATGTCCGAGCCCGCAGCGG 23780
QY 380 TCGTACAGGTGGGGCTCTTCGCCGTGCGCGCGTACGCGACACCCGCTTGCCCTCGGCTT 439
Db 23779 CCAGCATGCGCTGCCCTTGGTGGTGGCGCGATCCCGCACGGTCTTGCCGTCCGCCA 23720
QY 440 TGACCGAGTTACAGAGGGGAGCAGGTCCGTGGGATGCCCTCGACGTCCGCGCGGAC 498
Db 23719 CCAGCGCGGTACCAAGTTCCCGGATCTGTGCGCGCACCCCTTCGTACGATCCGCGGAA 23661

RESULT 16
ACC58250/c
ID ACC58250 standard; DNA; 31248 BP.
XX
AC ACC58250;
XX
DT 26-AUG-2003 (first entry)
XX
DE Coumermycin A1 biosynthetic gene cluster.
XX
KW Coumermycin A1; aminocoumarin; antibiotic; antibacterial; cytostatic;
KW gene; ds.
XX
OS Streptomyces rishiriensis.
XX
FH Key Location/Qualifiers
FT CDS 1..960
FT /*tag= a
FT /gene= "couG"
FT /note= "cuma", encodes ABR42475"
FT CDS 1297..1512
FT /*tag= b
FT /gene= "couY"
FT /note= "similar to Mycobacterium tuberculosis mbtH,
FT encodes ABR42476"
FT CDS 1818..3617
FT /*tag= c
FT /gene= "couH"
FT /product= "Acyl-CoA synthetase"
FT /note= "encodes ABR42477"
FT CDS 3621..4844
FT /*tag= d
FT /gene= "couJ"
FT /product= "Cytochrome P450"
FT /note= "cumd", encodes ABR42479"
FT CDS 5004..5780
FT /*tag= e
FT /gene= "couJ"
FT /product= "CouJ"
FT /note= "similar to 3-ketoacyl-(ACP)-reductase, encodes
FT ABR42479"
FT CDS 5777..6514
FT /*tag= f
FT /gene= "couK"
FT /product= "Putative reductase"
FT /note= "encodes ABR42480"
FT CDS 6521..8110
FT /*tag= g
FT /gene= "couL"
FT /product= "Putative ligase"
FT /note= "encodes ABR42481"
FT CDS 8124..8332
FT /*tag= h
FT /gene= "couM"
FT /product= "Glycosyltransferase"
FT /note= "encodes ABR42482"
FT CDS 9413..9700
FT /*tag= i
FT /gene= "couN1"
FT /product= "Unknown"
FT /note= "Cuml", encodes ABR42483"
FT CDS 9711..10778
```

```
FT      /*tag= j  
FT      /gene= "CouN2"  
FT      /product= "3-Ketoacyl-(ACP)-synthase"  
FT      /note= "CumJ, similar to Streptomyces peucetius Dpsc,  
FT      encodes ABR42484"  
FT      10775..11896  
FT      /tag= k  
FT      /gene= "CouN3"  
FT      /product= "Acyl-CoA-dehydrogenase"  
FT      /note= "encodes ABR42485"  
FT      12015..13520  
FT      /tag= l  
FT      /gene= "CouN4"  
FT      /product= "Acyl-CoA-synthetase"  
FT      /note= "proB, encodes ABR42486"  
FT      13524..13793  
FT      /tag= m  
FT      /gene= "CouN5"  
FT      /product= "CouN5"  
FT      /note= "similar to Pseudomonas fluorescens Pf-5 PtlL,  
FT      encodes ABR42487"  
FT      13851..15533  
FT      /tag= n  
FT      /gene= "CouN6"  
FT      /product= "Unknown"  
FT      /note= "cumK, encodes ABR42488"  
FT      15582..16427  
FT      /tag= o  
FT      /gene= "CouN7"  
FT      /product= "Putative hydrolase"  
FT      /note= "cumL, encodes ABR42489"  
FT      16777..17469  
FT      /tag= p  
FT      /gene= "CouO"  
FT      /product= "Putative methyltransferase"  
FT      /note= "cumM, encodes ABR42490"  
FT      17869..18699  
FT      /tag= q  
FT      /gene= "CouP"  
FT      /product= "O-methyltransferase"  
FT      /note= "cumN, encodes ABR42491"  
FT      18796..20220  
FT      /tag= x  
FT      /gene= "CouR1"  
FT      /product= "Unknown"  
FT      /note= "cumO, encodes ABR42492"  
FT      20220..21353  
FT      /tag= s  
FT      /gene= "CouR2"  
FT      /product= "Decarboxylase"  
FT      /note= "cumP, encodes ABR42493, similar to  
FT      phosphonopyruvate decarboxylase"  
FT      21386..22294  
FT      /tag= t  
FT      /gene= "CouR3"  
FT      /product= "CouR3"  
FT      /note= "cumO, encodes ABR42494, similar to Salmonella  
FT      enterica serovar typhimurium PdxU"  
FT      22349..223518  
FT      /tag= u  
FT      /gene= "CouR4"  
FT      /product= "Dehydrogenase"  
FT      /note= "cumS, encodes ABR42495, similar to lactate and  
FT      L(+)-mandelate dehydrogenases"  
FT      complement(23572..25047)  
FT      /tag= v  
FT      /gene= "CouR5"  
FT      /product= "Putative resistance protein"  
FT      /note= "cums, encodes ABR42496, transporter"  
FT      25185..26057  
FT      /tag= w  
FT      /gene= "CouR6"  
FT      /product= "transcript regulatory protein"
```

```
FT      /note= "cumT, encodes ABR42497"  
FT      26311..27177  
FT      /tag= x  
FT      /gene= "CouS"  
FT      /product= "dTDP-4-keto-6-deoxyhexose reductase"  
FT      /note= "cumU, encodes ABR42498"  
FT      27211..28221  
FT      /tag= y  
FT      /gene= "CouT"  
FT      /product= "dTDP-glucose 4,6-dehydratase"  
FT      /note= "cumV, encodes ABR42499"  
FT      28447..29709  
FT      /tag= z  
FT      /gene= "CouU"  
FT      /product= "C-methyltransferase"  
FT      /note= "cumW, encodes ABR42500"  
FT      29768..30658  
FT      /tag= aa  
FT      /gene= "CouV"  
FT      /product= "dTDP-glucose synthase; glucose-1-phosphate  
FT      thymidyltransferase"  
FT      /note= "cumX, encodes ABR42501"  
FT      30655..31248  
FT      /tag= ab  
FT      /gene= "CouW"  
FT      /product= "dTDP-4-keto-6-deoxyglucose 3,5-epimerase"  
FT      /note= "cumY, encodes ABR42502, no stop codon"  
XX      WO2003014352-A2.  
XX      20-FEB-2003.  
XX      06-AUG-2002; 2002WO-EP008777.  
XX      08-AUG-2001; 2001US-0310808P.  
XX      (UYTU-) UNIV TUEBINGEN.  
XX      Heide L, Li S;  
XX      WPI; 2003-268202/26.  
XX      P-PSDB; ABR42475, ABR42476, ABR42477, ABR42478, ABR42479, ABR42480,  
XX      ABR42481, ABR42482, ABR42483, ABR42484, ABR42485, ABR42486, ABR42487,  
XX      ABR42488, ABR42489, ABR42490, ABR42491, ABR42492, ABR42493, ABR42494,  
XX      ABR42495, ABR42496, ABR42497, ABR42498, ABR42499, ABR42500, ABR42501,  
XX      ABR42502.  
XX      Novel polypeptide encoded by open reading frame of aminocoumarin  
XX      biosynthetic gene cluster, useful for synthesizing aminocoumarins useful  
XX      for treating bacterial infections and malignancies.  
XX      Claim 2; Fig 1A; 124pp; English.  
XX      The present sequence is a partial sequence of the Streptomyces  
XX      rishiriensis strain DSM 40489 coumerycin A1 biosynthetic gene cluster.  
XX      The sequence was isolated from a cosmid library using probes based on  
XX      novobiocin-producer Streptomyces spheroides gene sequences. It includes  
XX      28 open reading frames (ORFs) that show marked sequence similarity to the  
XX      corresponding ORFs of the novobiocin cluster and in identical order. The  
XX      invention provides aminocoumarin (e.g. coumerycin A1) biosynthetic  
XX      nucleic acids and polypeptides, a method for producing hybrid antibiotics  
XX      using a combination of various regions of biosynthetic gene clusters, and  
XX      a pharmaceutical composition containing an aminocoumarin compound for use  
XX      in the treatment of Gram-positive bacterial infections and of malignant  
XX      diseases
```

Query Match 18.8%; Score 94; DB 8; Length 31248;  
Best Local Similarity 53.9%; Pred. No. 1.1e-08;  
Matches 193; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

Qy 140 TCAGTGTACGCGGGAGCTACAGGATCCACTGTCCGCCAGCCGCGGACTCTGCT 199

```
Db 29686 TCACGCGGTCGGGTATGGGAACATGAACAGCGCCGCCCTTCAGGAAACACTTCT 29627
QY 200 CTTTCGCCATGATCTCGTCGGCGTGTTCAGAGGAGAGAGAGCGCGTAGTCCACCGGT 259
Db 29626 CTTTCGGTATGATCTCCAGCGGTAGTTCCAGGCGCAACAGCAGGTAGTAGTCCGGTAGCT 29567
QY 260 CGGGCGTGACCGTCGGGTCGCGACCGGATGTGCGTCCGGGGTGAGCCGGCCT 319
Db 29566 GGTCCCTGCTGTTCGGCGACACAGATCGGACTCTCGAGCCCGTAGGAGTTTCCCT 29507
QY 320 GTTTCGCCGGGTCTGTGTCGACACCCAGGAGACAGGTTCGGACCCGATGCCGAGAAGT 379
Db 29506 GCTTCAGGCTCTTGTGTCGTCGACAGACCGGATCTCCTGATGCCAGCCCGATCGG 29447
QY 380 TCGTCACGTGCGCTCTTCGCGTCGCGCGTACCGCACCGCTTGCCTCGGCGCT 439
Db 29446 TCAGCAGTGCCTTGTGTCGTCGTCGCGTATCCAGCACACTGTCTTGGCGTCCCA 29387
QY 440 TGAGCGAGTTCAGCAGGCGGACAGGTCTGGTCCGATGCCCTCGACGTCCGCGCGAA 497
Db 29386 CCAAGTGGCGACCAAGTCCCGGATCTGTGCGCGCACCGCTTCCGTACGCTCGGCGAA 29329

RESULT 17
ACC58251/c
ID ACC58251 standard; DNA; 35359 BP.
XX
AC ACC58251;
XX
DT 26-AUG-2003 (first entry)
XX
DE Coumermycin A1 biosynthetic gene cluster.
XX
KW Coumermycin A1; aminocoumarin; antibiotic; antibacterial; cytostatic;
KW gene, ds.
XX
OS Streptomyces rishiriensis.
XX
FH Key Location/Qualifiers
CDS 1..1028
FT /*tag= a
FT /note= "ORF3, encodes ABR42503"
CDS complement(1045..1536)
FT /*tag= b
FT /product= "putative FK506-binding protein"
FT /note= "encodes ABR42504"
FT /*tag= c
FT /note= "ORF3, encodes ABR42505"
CDS 2306..2959
FT /*tag= d
FT /gene= "CouE"
FT /product= "CouE"
FT /note= "encodes ABR42506"
CDS 4112..5071
FT /*tag= e
FT /gene= "CouG"
FT /note= "cuma, encodes ABR42475"
CDS 5408..5623
FT /*tag= f
FT /gene= "CouY"
FT /note= "similar to Mycobacterium tuberculosis mbtH,
FT encodes ABR42476"
CDS 5929..7728
FT /*tag= g
FT /gene= "CouH"
FT /product= "Acyl-CoA synthetase"
FT /note= "encodes ABR42477"
CDS 7732..8955
FT /*tag= h
FT /gene= "CouI"
FT /product= "Cytochrome P450"
```

```
FT /*note= "cumD, encodes ABR42478"
FT 9115..9891
FT /*tag= i
FT /gene= "CouJ"
FT /product= "CouJ"
FT /note= "similar to 3-ketoacyl-(ACP)-reductase, encodes
FT ABR42479"
CDS 9888..10625
FT /*tag= j
FT /gene= "CouK"
FT /product= "Putative reductase"
FT /note= "encodes ABR42480"
CDS 10632..12221
FT /*tag= k
FT /gene= "CouL"
FT /product= "Putative ligase"
FT /note= "encodes ABR42481"
CDS 12235..13443
FT /*tag= l
FT /gene= "CouM"
FT /product= "Glycosyltransferase"
FT /note= "encodes ABR42482"
CDS 13524..13811
FT /*tag= m
FT /gene= "CouN1"
FT /product= "Unknown"
FT /note= "CumI, encodes ABR42483"
CDS 13822..14889
FT /*tag= n
FT /gene= "CouN2"
FT /product= "3-Ketoacyl-(ACP)-synthase"
FT /note= "CumJ, similar to Streptomyces peucetius DpsC,
FT encodes ABR42484"
CDS 14886..16007
FT /*tag= o
FT /gene= "CouN3"
FT /product= "Acyl-CoA-dehydrogenase"
FT /note= "encodes ABR42485"
CDS 16126..17631
FT /*tag= p
FT /gene= "CouN4"
FT /product= "Acyl-CoA-synthetase"
FT /note= "proB, encodes ABR42486"
CDS 17635..17904
FT /*tag= q
FT /gene= "CouN5"
FT /product= "CouN5"
FT /note= "similar to Pseudomonas fluorescens Pf-5 PtlL,
FT encodes ABR42487"
CDS 17962..19644
FT /*tag= r
FT /gene= "CouN6"
FT /product= "Unknown"
FT /note= "cumK, encodes ABR42488"
CDS 19693..20538
FT /*tag= s
FT /gene= "CouN7"
FT /product= "Putative hydrolase"
FT /note= "cumL, encodes ABR42489"
CDS 20888..21580
FT /*tag= t
FT /gene= "CouO"
FT /product= "Putative methyltransferase"
FT /note= "cumM, encodes ABR42490"
CDS 21980..22810
FT /*tag= u
FT /gene= "CouP"
FT /product= "O-methyltransferase"
FT /note= "cumN, encodes ABR42491"
CDS 22907..24331
FT /*tag= v
FT /gene= "CouR1"
FT /product= "Unknown"
```





PS Example 1; Page 273-274; 51pp; English.

XX The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 27-OCT-2003 to standardise OS field)

XX SQ Sequence 1290 BP; 209 A; 458 C; 441 G; 182 T; 0 U; 0 Other;  
Query Match 18.5%; Score 92.4; DB 7; Length 1290;  
Best Local Similarity 53.6%; Pred. No. 2.9e-08;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 142 AGGTGACCGGGGAGTACAGGATCCACTCTCCGCGACGCCCGCGGAATCTCTGCTCC 201  
DB 1250 ACCCGCAGGCTGGCGAGCGGAGATCAGCTTGTTCGGCGCGCCACCGTGCGCCCTCG 1191  
QY 202 TTCCGCATCATCTCGTGGCGGTGTTCAGGCGGAGAGCAGCGGTAGTTCACCGCGTGC 261  
DB 1190 CGGCGCACCATCTCTGTCGTGGAGTGCACGCGCAGCAGTGTGTCGGGCGGCG 1131  
QY 262 GCGGTGAACGTCGCGGGTGCACACCGGGATGTGCGTCCGGGGGTGAGCGGCCCTGC 321  
DB 1130 GCGCGGAGTGGCGCTCGTGTATCTCGATGTCCGTACCGAGGTCGCGCGCGGACC 1071  
QY 322 TTGGCGCGGTCTGTGCGCACCCAGGAGACAGTCCGCGACCGATCCGCGAGATTC 381  
DB 1070 TTGTCGGGTTCGGCTTCGGCGGCTACGGGATCAGGTGCGTTCGATCCCGAGTCTGC 1011  
QY 382 GTACCGTGGCGCTCTTCGCCGTGCGCGGTACGCCACACCCGCTTGCCTCGGCCCTTG 441  
DB 1010 AGGAGGCTGTGCGCTTGTGGAGGCGCGGTAGAGGTGACGCTGCTTCGCTTCGCGC 951  
QY 442 AGCGAGTTCAGAGGCGGAGCGGTGCGGTGCGGATGCCCTTCACGTCGCGCGGAACC 499  
DB 950 AGACCATGACGATCTTGACAGTTCGTGCGGTGCGCGCCGACGCTTGTGCGCGAACC 893

RESULT 19  
ABL50558/c  
ID ABL50558 standard; DNA; 14252 BP.  
XX AC ABL50558;  
XX 18-JUN-2002 (first entry)  
XX DE Micromonospora carbonacea everninomicin locus nucleotide contig 4.  
XX KW Micromonospora carbonacea; antibiotic; everninomicin; biosynthesis;  
XX gene cluster; genetic manipulation; contig; gene; ds.  
XX OS Micromonospora carbonacea.  
XX PN WO200155180-A2.  
XX PD 02-AUG-2001.  
XX 29-JAN-2001; 2001WO-CA000128.  
XX 27-JAN-2000; 2000US-0177711P.  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX (FARN/) FARNET C.

XX Staffa A, Zazopoulos E, Mercure S, Nowacki P;  
XX WPI; 2001-476185/51.  
DR P-PSDB; ABB06885, ABB06886, ABB06887, ABB06888, ABB06889, ABB06890,  
DR ABB06891, ABB06892, ABB06893, ABB06894, ABB06895, ABB06896, ABB06897.  
XX Novel isolated gene cluster encoding polypeptides involved in  
PT everninomicin biosynthesis useful for construction of everninomicin  
PT overproducing strains, and to allow chemical modifications of  
PT everninomicin to enhance certain properties.  
XX PS Claim 1; Fig 1; 181pp; English.  
XX ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora  
CC carbonacea everninomicin biosynthetic locus gene cluster. The contigs  
CC encode the protein sequences designated ORF (open reading frame) 1 to 49,  
CC given in ABB06881 to ABB06930. The gene cluster is useful for the  
CC construction of the everninomicin antibiotic in overproducing strains,  
CC and to allow chemical modifications of everninomicin to enhance certain  
CC properties via genetic manipulation or combinatorial biosynthesis. The  
CC gene cluster can be used to produce genetic systems and genes encoding  
CC novel enzyme activities, and avoid the problems of low yield and quality  
CC of everninomicins produced by chemical synthesis  
XX SQ Sequence 14252 BP; 2178 A; 4993 C; 4959 G; 2122 T; 0 U; 0 Other;  
Query Match 18.5%; Score 92.4; DB 4; Length 14252;  
Best Local Similarity 53.6%; Pred. No. 2.4e-08;  
Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
QY 142 AGGTGACCGGGGAGTACAGGATCCACTCTCCGCGACGCCCGCGGAATCTCTGCTCC 201  
DB 2529 ACCCGCAGGCTGGCGAGCGGAGATCAGCTTGTTCGGCGCGCCACCGTGCGCCCTCG 2470  
QY 202 TTGCCATGATCTCGTGGCGGTGTTCAGGCGGAGAGCAGCGGTAGTTCACCGCGTGC 261  
DB 2469 CGGCGCACCATCTCTGTCGTGGAGTGCACGCGCAGCAGGATGTTGTCGGGCGGCG 2410  
QY 262 GCGGTGAACGTCGCGGGTGCACACCGGGATGTGCGTCCGGGGGTGAGCGGCCCTGC 321  
DB 2409 GCGCGGAGTTCGCGCTCGCTGATCTCGATGTCCGTACCGAGGTCGCGCGCGGACC 2350  
QY 322 TTGCCCGCGCTGTGTCGCACACCCAGGAGACAGGTCCGCGACCGATCCGCGAGATTC 381  
DB 2349 TTGTCGGGTTCGCTTCGCGGCGGTACGGGATCAGCGTGCCTCCGCGAGTCTGC 2290  
QY 382 GTACCGTGGCGCTCTTCGCCGTGCGCGGTACGCCACACCCGCTTGCCTCGGCCCTTG 441  
DB 2289 AGGAGGCTGTGCGCTTGTGGAGGCGCGGTAGAGGTGACGCTGCTTCGCTGTCGCGC 2230  
QY 442 AGCGAGTTCAGAGGCGGAGCGGTGCGGTGCGGATGCCCTTCGCGCGCGGGAACC 499  
DB 2229 AGACCATGACGATCTTGACAGTTCGTGCGGTGCGCGCCGACGCTTGTGCGGGAACC 2172

RESULT 20  
ABZ66811/c  
ID ABZ66811 standard; DNA; 48221 BP.  
XX AC ABZ66811;  
XX 27-OCT-2003 (revised)  
XX 21-MAR-2003 (first entry)  
XX DE Orthosomycin biosynthetic gene cluster SEQ ID NO 280.  
XX KW Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX OS Micromonospora carbonacea; aurantiaca.  
XX PN WO200279505-A2.

PD 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-CA000432.  
XX  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-0279709P.  
XX 20-APR-2001; 2001US-0285214P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX WPI; 2003-058435/05.  
XX  
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene  
XX cluster, by detecting presence of nucleic acid sequence corresponding to  
XX 17 of flambamycins protein families.  
XX  
XX Example 1; Page 479-505; 511pp; English.  
XX  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of orthosomycin  
XX produced by the organism. This allows for the detection of new  
XX orthosomycin natural products, not produced by the organism. (Updated on  
XX 27-OCT-2003 to standardise OS field)  
XX  
XX Sequence 48221 BP; 7199 A; 16729 C; 16897 G; 7395 T; 0 U; 1 Other;  
XX  
XX Query Match 18.5%; Score 92.4; DB 7; Length 48221;  
XX Best Local Similarity 53.6%; Pred. No. 2.2e-08;  
XX Matches 192; Conservative 0; Mismatches 166; Indels 0; Gaps 0;  
XX  
XX 142 AGCTGTACGCGGGGACGTACAGGATCCACTGTCCGCGAGCCGCGGAACTCTCTGTCC 201  
XX  
XX 8244 ACCCGCAGGCTGGGACGGGAGATCAGCTTGTTCGCGCGCCACCGTGGCGCCTCG 8195  
XX  
XX 202 TTCGCATGATCTCTCGCGGTGTTCAGGCGAAGACAGCGGTAGTCCACCGCGTCG 261  
XX  
XX 8184 CGCGCCAGCATCTCGTGTGGAAGTGCACCGCAGCAGCAGTGTGTCTCGCGCGCGG 8125  
XX  
XX 262 GCGGTGAACGGCTCGCGGTGCGCACCGGATGTGCGTGGCGGTGAGCGCGCCCTGC 321  
XX  
XX 8124 GCCCGCGAGTCGCGCTCGCTGATGATCTCGATGTCCGTACCGAGGTCCGCGCGCG 8065  
XX  
XX 322 TTGCGCGCGGTGTGTGCGACACCCAGGAGACAGGTCCGACCGATCCCGCAGAGTTC 381  
XX  
XX 8064 TTGTCCGGTTCGCGTCCGCGCGGTACGGATACCGTGTGCGTCCGCGAGTACTGC 8005  
XX  
XX 382 GTACGGTGGCGCTTTCGCGTCCGCGGTACGCGACACCGTTCGCGTCCGCGCTTG 441  
XX  
XX 8004 AGGAGGGTGTTCGCTTGTGAGGCGCGGTAGAGTGCAGGTGCTTCGCGTTCGCG 7945  
XX  
XX 442 AGCGAGTTCAGCAGGCGCAGCAGGTTCGCTGCGGATGCGCTTCGCGCGCGCAACC 499  
XX  
XX 7944 AGACCATGAGCATCTTGACGAGTTTCGTGCGGTGCGCGCGCCGACGCTTTCGCG 7887  
XX  
XX  
XX RESULT 21  
XX ID ABZ66808/c  
XX ABZ66808 standard; DNA; 45055 BP.  
XX  
XX AC ABZ66808;  
XX  
XX DT 21-MAR-2003 (first entry)

XX Orthosomycin biosynthetic gene cluster SEQ ID NO 277.  
XX  
XX Orthosomycin; biosynthesis; everninomicin; avilamycin; gene; ds.  
XX  
XX Streptomyces mobaraensis.  
XX  
XX WO200279505-A2.  
XX  
XX 10-OCT-2002.  
XX  
XX 28-MAR-2002; 2002WO-CA000432.  
XX  
XX 28-MAR-2001; 2001US-0279095P.  
XX 30-MAR-2001; 2001US-0279709P.  
XX 20-APR-2001; 2001US-0285214P.  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX WPI; 2003-058435/05.  
XX  
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene  
XX cluster, by detecting presence of nucleic acid sequence corresponding to  
XX 17 of flambamycins protein families.  
XX  
XX Example 2; Page 425-450; 511pp; English.  
XX  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
XX ABP99362). The method is useful for identifying an orthosomycin  
XX biosynthetic gene, gene fragment or gene cluster, especially an  
XX everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
XX gene fragment or gene cluster. The method is useful for detecting the  
XX presence of any organism that contains DNA for the production of  
XX orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
XX orthosomycins) regardless of the level at which genes for orthosomycin  
XX production are expressed by the organism or the amount of orthosomycin  
XX produced by the organism. This allows for the detection of new  
XX orthosomycin natural products, not produced by the organism  
XX  
XX Sequence 45055 BP; 6228 A; 16576 C; 16043 G; 6208 T; 0 U; 0 Other;  
XX  
XX Query Match 18.1%; Score 90.6; DB 7; Length 45055;  
XX Best Local Similarity 52.5%; Pred. No. 4.6e-08;  
XX Matches 198; Conservative 0; Mismatches 179; Indels 0; Gaps 0;  
XX  
XX 122 CACCGCGGTGAGTCTCGTACCGTGTACCGGGGACGTACAGGATCCACTGTCCGCCAG 181  
XX 9974 CCCCGCACCGGTCAACCATCTTCAGGGGGGACGCGGAGACACAGTTTCCGCCCG 9915  
XX  
XX 182 CCGCGCGGAACTCTCTCTTTCGCGATGATCTGTCGCGGTGTTCAGGGGAGAGCA 241  
XX 9914 CCGCCAGGAAGTCCGCTTCCCGGGCGACCATCGCTCCGCTAGATCCACGGAGACCA 9855  
XX  
XX 242 GCGCGTAGTCCACCGGTTCGGGTGAACCGCTCCGGGTGCGCACCGGGATGCGTGC 301  
XX 9854 GGAACCGGTTCGCGTCCAGCGCCCGGCGCTCTTCTCGGAGAGATCGGATGTCGTC 9795  
XX  
XX 302 CCGCGGTGAGCGCGCGCTCTGTCGCGCGGTCTGTGTGCGACACCCAGGAGACAGTCCG 361  
XX 9794 CCGCGGTGAGCGAGCGAGCTTCACGTTCGCTCTTTCACCTCGCGGATGACCGGAGCGTCT 9735  
XX  
XX 362 GACCGATGCGCGAGAGTTCGTTCACGTTCGCGCTCTTCGCGGTTCGCGGTAGCGCACCA 421  
XX 9734 CGTCCAGGCGCAGTACTGAGGATACGCTTCCGCTTGTGGAGGCGCGCTACCGGAGG 9675  
XX  
XX 422 CCGCGTTCGCGCTTCGCGCTTTCAGCGAGTTCAGCAGGCGCGAGCGTTCGCGGATCCCT 481  
XX 9674 TCGCGACCCCTTGTTCGCGGAGGTGCTGAGGAGTCCACGAGCCGCTCGCGGATTCCT 9615



Db 44857 AGGCCAGGAAGTCGGTTCCTCCGGCGACCAATCGCTCCCGGTAGATCCACGGGAGGACCA 44916  
 QY 242 CGCGGTAGTCCACCGCGTCGGCGGTGAACGGGTCCGGGTGCGCACCGGGATGTGCGTGC 301  
 Db 44917 GGAAGTGTGGGTCAAGCGCCCGGGCTCTCTCGGAGACGATCGGATGTTCGTGC 44976  
 QY 302 CGGGGTGAGCCGGCCCTGCTTGGCGCGGTGCTGTGCACACCCAGGAGACAGGTCG 361  
 Db 44977 CGGGCGTGTAGACGCGCAACTTGTCTCTGTTCACTCGGGATGACGGGAGGAGCGTCT 45036  
 QY 362 GACCGATGCCCGAGAAGTTCGTACGGTGGGGCTTTCGCGGTGCGCGCGGTACGCCACCA 421  
 Db 45037 CGTCCAGCCCGAGTACTGCAAGGATGACGTGCGCTTCGTGAGCGCGCGGTACCGAGG 45096  
 QY 422 CCCGCTTGGCCCTCGGCGCTTGAAGTTCAGCAGGGCGAGAGTCGCGTGGGATGCGCT 481  
 Db 45097 TGTGAGCGCCCTTGTCCCGCAGCGGTGAGGAAATTCAGGAGTCGCTACGGGACTTCT 45156  
 QY 482 CGACGTGGCGGCGGAAC 498  
 Db 45157 CGTCCCGCGGCGGAAC 45173  
 RESULT 24  
 ABZ37515/C  
 ID ABZ37515 standard; DNA; 59816 BP.  
 XX AC ABZ37515;  
 XX DT 05-MAR-2003 (first entry)  
 XX DE Streptomyces viridochromogenes Avi gene cluster sense strand.  
 XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; gene; ds.  
 XX OS Streptomyces viridochromogenes.  
 XX PN WO200268436-A1.  
 XX PD 06-SEP-2002.  
 XX PF 24-AUG-2001; 2001WO-EP009815.  
 XX PR 25-FEB-2001; 2001DE-01009166.  
 XX PA (COMB-) COMBINATURE BIOPHARM AG.  
 XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX DR WPI; 2003-018650/01.  
 XX DR P-PSDB; ABB98398, ABP76679.  
 XX PT New avilamycin derivatives, useful for treatment of infections, and  
 XX PT nucleic acid encoding avilamycin synthesis enzymes.  
 XX PS Example 1; Page 68-301; 319pp; German.  
 XX CC The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of a nucleic acid encoding avilamycin synthesis enzymes  
 CC from the Streptomyces viridochromogenes Avilamycin A biosynthetic gene  
 CC cluster (ABZ37515-ABZ37516)  
 XX SQ Sequence 59816 BP; 8915 A; 21088 C; 21053 G; 8760 T; 0 U; 0 Other;

Query Match 16.2%; Score 81; DB 7; Length 59816;  
 Best Local Similarity 50.9%; Pred. No. 2.5e-06;  
 Matches 192; Conservative 0; Mismatches 185; Indels 0; Gaps 0;

QY 122 CACCGCGGGTCAGTCTGTCAGTGTACGCGGGGAGCGTACAGGATCCACTCTCGCGCCAG 181  
 Db 15020 CTTCGCGCAACCGATCACACAATTCAGCGTGGCGAGCGGAAAGACCAAGGCTGCTCCGG 14961  
 QY 182 CCGCGGGGAACCTCTGCTCTTCGCAATGATCTCGTCGCGGTGGTTCAGGGCGAAGACCA 241  
 Db 14960 AGGCCAGGAAGTTCGGTTCCTCCGGGCGACCATCGCTCCGGGTAGATCCACGGGAGACCA 14901  
 QY 242 GCGCGTAGTCCACCGCGTGGCGGTGAACGGCTCCGGGGTGCACCGGATGTGCGTGC 301  
 Db 14900 GGAACCTGTGTCGGCTCAAGCGCCCGGGCTCTCTCGGAGACGATCGGATGTTCGTGC 14841  
 QY 302 CGGGGTGAGCGCGCGCTCTTGGCGCGGTGCTGCGCACACCCAGGAGACCAAGTCCG 361  
 Db 14840 CGGGCGTGTAGCAGCGCAACTTGTCTCTGTTCACTCGCGATGACGGCAGGAGCTCT 14781  
 QY 362 GACCGATGCCCGAGAAGTTCGTACGGTGGCGCTCTTCGCGGTGCGCGGTACGCCACCA 421  
 Db 14780 CGTCCAGCGCGCAGTACTGCAAGGATGACGTTGCTTCTGCGAGCGCGGTAGCCAAGG 14721  
 QY 422 CGCGCTTGGCCCTCGGCGTTCAGCGAGTTCAGCAGGGCGAGCGTGGTGGGATGCCCT 481  
 Db 14720 TGTGAGCGCCCTTGTCCCGCAGCGGTGAGGAAATTCAGGAGTCGCTACGGGACTCTCT 14661  
 QY 482 CGACGTGGCGGCGGAAC 498  
 Db 14660 CGTCCCGCGGCGGAAC 14644  
 RESULT 25  
 ABZ37534/C  
 ID ABZ37534 standard; DNA; 1224 BP.  
 XX AC ABZ37534;  
 XX DT 26-FEB-2003 (first entry)  
 XX DE Streptomyces viridochromogenes AviI encoding polynucleotide.  
 XX KW Avilamycin; antibacterial; virucide; protozoacide; fungicide; infection;  
 XX KW medicine; Staphylococcus aureus; biosynthetic gene cluster; gene; ds.  
 XX OS Streptomyces viridochromogenes.  
 XX PN WO200268436-A1.  
 XX PD 06-SEP-2002.  
 XX PF 24-AUG-2001; 2001WO-EP009815.  
 XX PR 25-FEB-2001; 2001DE-01009166.  
 XX PA (COMB-) COMBINATURE BIOPHARM AG.  
 XX PI Weitnauer G, Muehlenweg A, Trefzer A, Bechthold A;  
 XX DR WPI; 2003-018650/01.  
 XX DR P-PSDB; ABP76700.  
 XX PT New avilamycin derivatives, useful for treatment of infections, and  
 XX PT nucleic acid encoding avilamycin synthesis enzymes.  
 XX PS Claim 13; Page 121-126; 319pp; German.  
 XX CC The invention relates to avilamycin derivatives (I) with antibacterial,  
 CC virucide, protozoacide and fungicide activity. (I) are useful for  
 CC treatment of infections (bacterial, viral, protozoal or fungal), in human  
 CC or veterinary medicine, particularly where caused by Staphylococcus  
 CC aureus. (I) are more hydrophilic than known avilamycins. The present  
 CC sequence is that of a nucleic acid encoding avilamycin synthesis enzymes  
 CC from the Streptomyces viridochromogenes Avilamycin A biosynthetic gene  
 CC cluster (ABZ37515-ABZ37516)

XX SQ Sequence 1224 BP; 179 A; 456 C; 396 G; 193 T; 0 U; 0 Other;  
Query Match 16.2%; Score 80.8; DB 7; Length 1224;  
Best Local Similarity 53.1%; Pred. No. 3.8e-06;  
Matches 172; Conservative 0; Mismatches 152; Indels 0; Gaps 0;  
QY 175 CCACAGCCGGCGAATCTCTCTCTCGCCATGATCTCGCGCGTGTTCAGGGCG 234  
Db 1184 CCTCCGAGGCCAGGAAGTCGGTTCCTCCGGCGACCATCGCGTAGATCCACGGG 1125  
QY 235 AAGAGAGCCGCTAGTCCACCGCGTCGGGGTGACCGTCCGGGGTGCACCGGGATG 294  
Db 1124 AGACAGGAACCTGGTGGGCTCAAGCGCCCGGCTCTCTCGGAGACGATCGGGATG 1065  
QY 295 TCGTCCCGGGGTGAGCGCCCTGCTTGGCCGGTGTGTCGACACCCAGAGACC 354  
Db 1064 TTGTCGCCGGCTGTAGACCGAAGTCTCTCTCTTACCTCGCGATGAGGGCAGG 1005  
QY 355 AGTTCGGACCGATGCGCAGAAAGTTCGTCAAGTGGCGCTTTCCTCGCGCGGTAC 414  
Db 1004 AGCGTCTGTCAGCGCGCAGTACTGACGATGACCTTGCCTTCGTGAGGCGCGGTAG 945  
QY 415 GCCACACCGCTTGCCTCGGCTTGAGCGAGTTCAGAGGCGGAGCGAGTTCGGTGG 474  
Db 944 CCAGGGTGTGACGCCCTTGTCCCGGAGGCGGTGAGGAATTCCAGGAGTCTGTCACGG 885  
QY 475 ATGCCCTCGACGTCGGCGCGAAC 498  
Db 884 GACTCCTCGTCCGCGCGCGAAC 861  
RESULT 26  
ID ACC58253/c  
XX AC ACC58253;  
XX DT 27-OCT-2003 (revised)  
XX DT 26-AUG-2003 (first entry)  
XX DE Clorobiocin biosynthetic gene cluster.  
XX KW Clorobiocin; aminocoumarin; antibiotic; antibacterial; cytostatic; gene;  
XX ds.  
XX OS Streptomyces roseochromogenes subsp. oscitans.  
FH Key Location/Qualifiers  
FT CDS 18..938  
FT /note= "ORF1, encodes ABR42509, similar to Streptomyces  
FT hygroscopicus FxbL"  
FT /tag= b 1613  
FT /product= "sarcosine oxidase gamma subunit"  
FT /note= "ORF2, encodes ABR42510"  
FT complement(1642..4530)  
FT /tag= c  
FT /product= "sarcosine oxidase alpha subunit"  
FT /note= "ORF3, encodes ABR42511"  
FT complement(4527..4808)  
FT /tag= d  
FT /product= "sarcosine oxidase delta subunit"  
FT /note= "ORF4, encodes ABR42512"  
FT complement(4819..6039)  
FT /tag= e  
FT /product= "sarcosine oxidase beta subunit"  
FT /note= "ORF5, encodes ABR42513"  
FT complement(6036..7256)  
FT /tag= f  
FT /product= "Unknown"  
FT /note= "ORF6, encodes ABR42514, similar to serine

hydroxymethyltransferase"  
7463..8119  
/tag= g "Unknown"  
/product= "Hypothetical protein"  
/note= "ORF7, encodes ABR42515, similar to gnt-R family  
transcriptional regulator"  
8265..8714  
/tag= h  
/product= "Hypothetical protein"  
/note= "ORF8, encodes ABR42516"  
8963..9199  
/tag= i  
/product= "Putative transposase"  
/note= "ORF9, encodes ABR42517"  
9348..10001  
/tag= j  
/gene= "CioE"  
/product= "CioE"  
/note= "encodes ABR42518, similar to novE of novobiocin  
biosynthetic gene cluster"  
10078..11166  
/tag= k  
/gene= "CioF"  
/product= "Putative oxidoreductase"  
/note= "encodes ABR42519, similar to novP of novobiocin  
biosynthetic gene cluster"  
11746..12705  
/tag= l  
/gene= "CioG"  
/product= "Regulatory protein clog"  
/note= "encodes ABR42520"  
12920..13135  
/tag= m  
/gene= "CioY"  
/product= "mbtH-like protein"  
/note= "encodes ABR42521"  
13390..15192  
/tag= n  
/gene= "CioH"  
/product= "Peptide synthetase-like protein"  
/note= "encodes ABR42522"  
15196..16419  
/tag= o  
/gene= "CioI"  
/product= "Cytochrome P450"  
/note= "encodes ABR42523"  
16573..17349  
/tag= p  
/gene= "CioJ"  
/product= "Putative 3-oxoacyl-(ACP)-reductase"  
/note= "encodes ABR42524"  
17346..18083  
/tag= q  
/gene= "CioK"  
/product= "Putative oxidoreductase"  
/note= "encodes ABR42525"  
18090..19673  
/tag= r  
/gene= "CioL"  
/product= "Amide synthetase"  
/note= "similar to novobiocin acid synthetase, encodes  
ABR42526"  
19687..20859  
/tag= s  
/gene= "CioM"  
/product= "Glycosyltransferase"  
/note= "encodes ABR42527"  
20853..21140  
/tag= t  
/gene= "CioN"  
/product= "Hypothetical protein"  
/note= "encodes ABR42528"  
21150..22217

FT	/tag= u	
FT	/gene= "clON2"	
FT	/product= "Putative 3-ketoacyl- (ACP)- synthetase"	
FT	/note= "encodes ABR42529"	
CDS	22207..23337	
FT	/tag= u	
FT	/gene= "clON3"	
FT	/product= "Putative acyl-CoA dehydrogenase"	
FT	/note= "encodes ABR42530"	
CDS	23456..24961	
FT	/tag= v	
FT	/gene= "clON4"	
FT	/product= "Acyl-CoA synthetase"	
FT	/note= "encodes ABR42531"	
CDS	24965..25234	
FT	/tag= w	
FT	/gene= "clON5 /product= "ClON5methyltransferase"	
FT	/note= "encodes ABR42532, similar to acyl carrier protein"	
CDS	25292..26977	
FT	/tag= x	
FT	/gene= "clON6 /product= "ClON6"	
FT	/note= "encodes ABR42533, similar to couN6 of coumermycin A1"	
CDS	27026..27862	
FT	/tag= y	
FT	/gene= "clON7"	
FT	/product= "Putative hydrolase"	
FT	/note= "encodes ABR42534"	
CDS	28099..29673	
FT	/tag= z	
FT	/gene= "clOHal"	
FT	/product= "Putative halogenase"	
FT	/note= "encodes ABR42535"	
CDS	29998..30831	
FT	/tag= aa	
FT	/gene= "clOP"	
FT	/product= "O-methyltransferase"	
FT	/note= "encodes ABR42536"	
CDS	31832..32806	
FT	/tag= ab	
FT	/gene= "clOQ"	
FT	/product= "ClOQ"	
FT	/note= "encodes ABR42537, similar to novQ of novobiocin biosynthetic gene cluster"	
CDS	32803..33636	
FT	/tag= ac	
FT	/gene= "clOR"	
FT	/product= "Putative aldolase"	
FT	/note= "encodes ABR42538"	
CDS	34466..35332	
FT	/tag= ad	
FT	/gene= "clOS"	
FT	/product= "dTDP-4-keto-6-deoxyhexose reductase"	
FT	/note= "encodes ABR42539"	
CDS	35366..36376	
FT	/tag= ae	
FT	/gene= "clOT"	
FT	/product= "dTDP-glucose 4,6-dehydratase"	
FT	/note= "encodes ABR42540"	
CDS	36602..37864	
FT	/tag= af	
FT	/gene= "clOU"	
FT	/product= "C-methyltransferase"	
FT	/note= "encodes ABR42541"	
CDS	37976..38866	
FT	/tag= ag	
FT	/gene= "clOV"	
FT	/product= "dTDP-glucose synthase"	
FT	/note= "encodes ABR42542"	
CDS	38863..39459	
FT	/tag= ah	
FT	/gene= "clOW"	
FT	/tag= ai	

FT	/product= "dUDP-4-keto-6-deoxyglucose 3,5-epimerase"
FT	/note= "encodes ABR42543"
FT	39459. .40220
CDS	/*tag= ai
FT	/gene= "cldZ"
FT	/product= "Unknown"
FT	/note= "encodes ABR42544"
CDS	40574. .42291
FT	/*tag= aj
FT	/gene= "gyrBr"
FT	/partial
FT	/product= "Chlorobiotin-resistance gyrase B"
FT	/note= "encodes ABR42545, no stop codon"
XX	
PN	WO2003014352-A2.
Query Match	
Best Local Similarity 15.6%; Score 78; DB 8; Length 42291;	
Matches 183; Conservative 0; Mismatches 175; Indels 0; Gaps 0;	
QY	140 TCACGTGTCACGGGGAGCTACAGATCCACTTCCGCCACCGCGGGCGGAATCCTTGCT 199
Db	37841 TCACCGCTGGTTCCGGGTATGGAAACAATAAACGCGCCGCTTTTCCAGAACAACACTTCT 37782
QY	200 CTTTCGCATGATCTCGTCGGCGTGTTCCAGGCGGAGAGCAGCGCGTAGTCCACC CGGT 259
Db	37781 CTTTCTGATGATCTCCGGCGGTAGTTCCAGGCCNACAGAGGTAGTAGTCCGGTACAT 37722
QY	260 CGGGCGTGAA CGCGTCCGGGTGCGCACCGGGATGTGGCTCCGGGGGTGAGCGGGCCT 319
Db	37721 GGTCGCTGCTGTTTCCGGTACCAAGATCGGAAATTCTGGACCTCTGGTAGGATTTTTCCCT 37662
QY	320 GTTTCGGCGCGCTCGTGTGCGACACCCAGGAGAACAGGTCCGGGACCGATGCGCGAGAAGT 379
Db	37661 GTTTCAGAGCTGTTGTGTCACTACAGAACCGGATCTCTGATGCCCCAGCGCATGCGG 37602
QY	380 TCGTACGGTGGCGCTCTTCGGCGTCGGCGGTAGCGCAACACCGCGCTTGCCCTCGGCT 439
Db	37601 CGAGCAGTGCAC TGCGCTTTAGTTGGTGGCGCGGTATCCAGCACTGACTTGGCCCTTCGCA 37542
QY	440 TGAGCGAGTTTCAGAGGGCGAGCAGGTGCGGTGCGGATGCCCTCGACGTCGCGCGCAA 497
Db	37541 CCACGCGCATACCAAGTTCTCGGATCTGTGCGCAACTCGTTCCGTACGCTCGGCGAA 37484

RESULT 27  
ABZ66778/c  
ID ABZ66778 standard: DNA: 624 BP.

XX	
DT	27-OCT-2003 (revised)
DT	21-MAR-2003 (first entry)

XX Orthosomycin biosynthetic polynucleotide SEQ ID NO 218.  
DE  
XX  
KW Orthosomycin: biosynthesis; everninomicin; avilamycin; gene: ds.

XX OS Micromonospora carbonacea: aurantiaca.

XX PW WO200279505-A2.

XX  
PD 10-OCT-2002.

XX  
PF 28-MAR-2002; 2002WO-CA000432.

AA  
PR 28-MAR-2001; 2001US-0279095P.

PR 20-APR-2001; 2001US-0285214P.  
PR 20-APR-2001; 2001US-0285214P.

PA (ECOP-) ECOPIA BIOSCIENCES IN

PI Farnet CM, Zazopoulos E, Staf

XX WPI: 2003-058435/05.  
DR P-PSDB; ABP99315.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
XX Claim 16; Page 368-369; 511pp; English.  
PS  
XX The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ66670-ABZ66683), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 27-OCT-2003 to standardise OS field)  
XX  
XX Sequence 624 BP; 83 A; 225 C; 201 G; 115 T; 0 U; 0 Other;  
PS  
Query Match 11.9%; Score 59.4; DB 7; Length 624;  
Best Local Similarity 71.6%; Pred. No. 0.032;  
Matches 78; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 1 GGTACCGGACCGGTCCGGAACACGAGTCGAGATACGCGGAGGAGACACCCCGGT 60  
DB 127 GATATCGACGCTCTGGTGAACACCGAGTCGAGTACGCGGACGAGACACCCCGG 68  
QY 61 AGTCGGGTAGACGGTGGCGCGGAGCGGTACCGGCTTCGACGGTCAG 109  
DB 67 GGTCCGGAAGACCCCGGGGTGAAGACGTACCGCGCCCTCGACGGCGAG 19  
RESULT 28  
ID AAT70152/c  
XX AAT70152 standard; DNA; 2122 BP.  
XX AC AAT70152;  
XX 17-OCT-2003 (revised)  
XX 23-JUL-1997 (first entry)  
XX  
XX S.longisporoflavus staurosporin synthesis gene cluster 2.1kb fragment.  
XX  
XX Staurosporin; gene cluster; indole-carbazole alkaloid; ICA; antibiotic;  
XX antiproliferative; platelet aggregation; fungus; yeast; calcium ion;  
XX Ca2+/phospholipid-dependent serine/threonine protein kinase; ds.  
XX  
XX Streptomyces longisporoflavus; (strain R19).  
XX  
XX Key Location/Qualifiers  
FH 148..768  
CDS /\*tag= a  
FT /label= Gene 2  
FT /note= "Encodes a protein containing 207 amino acids  
FT significantly similar to known S-adenosyl methionine-  
FT dependent methyl transferases especially those of  
FT Streptomyces and Actinomycetes"  
FT 844..1686  
CDS /\*tag= b  
FT /label= Gene 1  
FT /note= "Encodes a protein containing 280 amino acids  
FT significantly similar to the dUDP-4-keto-6- deoxyglucose  
FT 3,5-epimerase of Streptomyces glaucoscescens i.e. has 48.6%  
FT homology"  
FT  
FT

FN WO9708323-A1.  
XX  
XX PD 06-MAR-1997.  
XX  
XX 19-AUG-1996; 96WO-EP003643.  
PF  
XX 30-AUG-1995; 95EP-00810534.  
PR  
XX (CIBA ) CIBA GEIGY AG.  
PA  
XX Schupp T, Engel N, Bietenhader J, Toupet C, Pospiech A;  
PI WPI; 1997-179280/16.  
XX P-PSDB; AAW17977; AAW17978.  
DR  
XX Indole-carbazole alkaloid biosynthesis gene cluster - especially coding  
PT for the antibiotic staurosporin from Streptomyces longisporoflavus.  
FT  
XX Claim 7; Page 32-33; 55pp; English.  
PS  
XX The present sequence represents the 2.1kb BglII fragment of Streptomyces  
CC longisporoflavus R19 which is involved in the biosynthesis of indole-  
CC carbazole alkaloids (ICA). The sequence contains two functional fragments  
CC as indicated in the features table. The DNA or a hybrid vector containing  
CC it can be used to prepare an ICA or derivatives and precursors, either by  
CC allowing production in previously incapable organisms or by improving  
CC yields. In particular, the antibiotic staurosporin can be produced.  
CC Staurosporin is known to have inhibitory activity against fungi, yeasts,  
CC and Ca2+/phospholipid-dependent serine/threonine protein kinases (PKCs).  
CC Staurosporin also has antiproliferative activity and can inhibit platelet  
CC aggregation. The present sequence can also inactivate ICA biosynthesis  
CC genes and can be used in PCR amplification. An advantage of this is that  
CC productivity of staurosporin-synthesising Streptomyces is improved over  
CC natural strains yielding only low concentrations. (Updated on 17-OCT-2003  
CC to standardise OS field)  
XX  
XX Sequence 2122 BP; 334 A; 803 C; 670 G; 315 T; 0 U; 0 Other;  
SQ  
Query Match 11.4%; Score 56.8; DB 2; Length 2122;  
Best Local Similarity 66.1%; Pred. No. 0.087;  
Matches 82; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 1 GGTACCGGACCGGTCTCCGGAACACGAGTCGAGATACGCGGAGGAGACACCCCGGT 60  
DB 271 GGGTCCCGTGGGCGCTCGGTGAACGTCTTCTCGTACGGCGGACGAACTTGCCTCTGT 212  
QY 61 AGTCGGGTAGACGGTGGCGCGGAGCGGTACGCGGCTTCGACGGTCAGCGGCGGCGG 120  
DB 211 GGTCCGGAAGACCGCGGGGTGAACGCGGCTCCTCGACGGTGAGCGGGGTGTT 152  
QY 121 ACAC 124  
DB 151 TCAC 148  
RESULT 29  
AAC85191  
ID AAC85191 standard; DNA; 9994 BP.  
XX  
XX AC AAC85191;  
XX 14-MAY-2001 (first entry)  
DT  
XX S. avermitilis 10 kb genomic DNA.  
DE  
XX Glycosylation; avermectin; aglycone; biocatalysis; antibiotic;  
XX cleandrose; ds.  
XX  
XX Streptomyces avermitilis.  
OS  
XX Key Location/Qualifiers  
FH 508..1332  
CDS /\*tag= a  
FT





antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pt\_sequences

XX SQ Sequence 837 BP; 158 A; 290 C; 264 G; 125 T; 0 U; 0 Other;

Query Match 11.0%; Score 54.8; DB 7; Length 837;  
Best Local Similarity 46.2%; Pred. No. 0.22;  
Matches 182; Conservative 0; Mismatches 212; Indels 0; Gaps 0;  
QY 106 TCAGCGGGGGGGACACCGCGGTGTCAGCTGTCACGTGACCGGGGGACGTACACAGG 165  
DB 539 TCAGCTGGGTGGTGGCCACGATGCCGTATTTCGGGTGCTTCGGGGGACCTCCTGT 480  
QY 166 ATCCATGTCGCGCACCGCGGGAACTCCTGTCCTTCGCCATGATCTGTCGCGGTGG 225  
DB 479 ACGCCAGACAGCTGCGCTTGGCGCGCCACCGCCATCGATCAGCTGCTTGAGCACCGC 420  
QY 226 TTCACGCGAAGACGACGCGTAGTCCACCGGTGCGGGGTGAACGCGTCCGGGTGCGC 285  
DB 419 GTCTCGGTCGATCAGGTATCGGACGACGACGCGAAGGGTCTGTGCTACCGCC 360  
QY 286 ACCGGGATGTCGTGCGGGGGTGAAGCGGCGCTGTCGCCGGGTGTCGACACAC 345  
DB 359 GCGCGGCGCTCAGCACGCGCATGTCCAGGCGCCAGCGGGCGCATTTGGCGGATGTACAG 300  
QY 346 CAGGACACAGGTCCGACCGATGCGCAGAGTTCGTACCGTGGCGCTCTTCGCGCTC 405  
DB 299 CAGTTCAGGTGCGCGGGGAGATGTGCGGACCGAGCCGACGATGTCGTGTCCTTC 240  
QY 406 GCGCCGTACGCCACACCGCTTGCCTTCGGCTTACGCGAGTTACAGGCGCGACGACG 465  
DB 239 GCTTCCAGGTGCGTTTCCAGCTCGGGGGCGGATCGAATGGTCTTCGATGCGCGCTTG 180  
QY 466 TCGGTGGGATGCCCTTCGAGCTCGGGGGGAC 499  
DB 179 TTGCGGCGCGGTGACGAAGATGAGGTGCGGTATCC 146

RESULT 31

AAI61173/c  
ID AAI61173 standard; DNA; 9975 BP.

XX AC AAI61173;

XX DT 22-SEP-2003 (first entry)

XX DE Actinosynnema pretiosum polyketide synthase (PKS) gene #4.

XX KW Maytansinoid; ansamitocin; antitumor; polyketide synthase; enzyme; PKS;

XX RW Gene; ds.

XX AC AAI61224;

OS Actinosynnema pretiosum.

XX PN WO2003045312-A2.

XX PD 05-JUN-2003.

XX PF 21-NOV-2002; 2002WO-US037547.

XX PR 21-NOV-2001; 2001US-0332158P.

XX PA (UNIW ) UNIV WASHINGTON.

XX PI Floss HG, Yu T, Leistner E;

XX DR WPI; 2003-493374/46.

XX PT Novel maytansinoid produced by bacterial host cell transformed with expression vector comprising open reading frame from ansamitocin gene cluster I of Actinosynnema pretiosum, useful as antitumor agent.

XX PS Disclosure; Page 62-67; 160pp; English.

XX CC The invention relates to maytansinoid produced by bacterial host cell transformed with expression vector comprising open reading frame from ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is useful as a potent antitumor agent. The present sequence is A. pretiosum ansamitocin gene cluster I polyketide synthase (PKS) gene

XX SQ Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;

Query Match 10.9%; Score 54.6; DB 7; Length 9975;  
Best Local Similarity 48.7%; Pred. No. 0.19;  
Matches 210; Conservative 0; Mismatches 214; Indels 7; Gaps 2;

QY 62 GTCGGGTACAGTGGCGCGAAGCGGTACGGCTTCACGCTCAGCGGGGGCGGCGGA 121  
DB 4227 GGCAGGTGACACGCCGTGAGCGGTGCGCGGGGACCG-CGTGAGCGCGCGG 4169  
QY 122 CACCGCGGGTCAAGTCTGTCACTGACGCGGGGACGTACAGGATCCATGTCGCGCAG 181  
DB 4168 CGAGGGCGCGCGGTGCGCGACGTGACGCGCGGGACACACCGCGGGCGCGCGCG 4109  
QY 182 CCGCGCGGAATCTCTGCTCTTCGCCATGATCTGCGGGGTGTTCCAGGCGAAGCA 241  
DB 4108 TCAGTCCGCGACACAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4049  
QY 242 GCGCGTAGTCCACCGCGTCCGGCGGTGAACCGCTCCGGGGTCCGACCCGCGATGTGCGTGC 301  
DB 4048 GCAGTGGCGCGCGCGCGCGGTGACGAGGTGGCGGGCGACCGCGCGCGCGCGCGCG 3989  
QY 302 CCGGGGTGAGCGCGCGCTGCTTGGCGCGCGTGTGCGACACCCAGGACACCGAGTCCG 361  
DB 3988 CGGTGCGCGCGGTACACGAGACGCTGCGGGTCCAGGACGGGTGCGGGCGCGCTCGG 3929  
QY 362 GACCGATGCGCGCAGAACTGCTACCGGTGGCGCTCTTCGCCGTGCGCGCGCGCGCA 421  
DB 3928 GGGCGCGCGCGCGCAACCTCGCGCGGTGACCGCGCGCGCGCGCGCGCGCGCG 3875  
QY 422 CCCGCTGCGCTTCGGCGGTGAGCGAGTTACAGCGGGGACGAGTCCGTCGGGATGCCCT 481  
DB 3874 GGGGCTCCGGTCCGGCGCGAGCGCGGTGATCAGGTCTGTGCGGGGCGGAGCGCGCTGT 3815  
QY 482 CGAGTGGCGG 492  
DB 3814 CCACGAGGACG 3804

RESULT 32

AAI61224

ID AAI61224 standard; DNA; 82746 BP.

XX AC AAI61224;

DT 22-SEP-2003 (first entry)  
XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX Actinosynnema pretiosum.  
XX WO2003045312-A2.  
XX 05-JUN-2003.  
XX 21-NOV-2002; 2002WO-US037547.  
XX 21-NOV-2001; 2001US-0332158P.  
XX (UNIW ) UNIV WASHINGTON.  
XX Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
XX expression vector comprising open reading frame from ansamitocin gene  
XX cluster I of Actinosynnema pretiosum, useful as antitumour agent.  
XX Claim 7; Page 105-152; 160pp; English.  
XX The invention relates to maytansinoid produced by bacterial host cell  
XX transformed with expression vector comprising open reading frame from  
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
XX useful as a potent antitumour agent. The present sequence is A. pretiosum  
XX ansamitocin biosynthetic gene cluster I  
XX  
SQ Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;  
Query Match 10.9%; Score 54.6; DB 7; Length 82746;  
Best Local Similarity 48.7%; Pred. No. 0.16;  
Matches 210; Conservative 0; Mismatches 214; Indels 7; Gaps 2;  
QY 62 GTCCGGTAGACGCTGCGCGCAAGGCTACCGCCTTCAGCTGACGGCGCGGCGGA 121  
Db 62974 GACCAGGTGTCACGACCGCGGTGAGCGGGTGCAGCGCGGGGACCG-CGTGAGCGCGCGG 63032  
QY 122 CACCGCGGCTCAGCTGCTACGCTACGCGGGGACGTACAGGATCCACTGTCCGCGAG 181  
Db 63033 CGAGGCGCGCGGCTGCGCGACGTGCGCGCGGACACACCGCGGCGCGCGCGCGG 63092  
QY 182 CCCGCGGAATCT 241  
Db 63093 TCAGCTCCGCGACAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 63152  
QY 242 GCGCGTAGTCCACCGCGTCCGCGGTGACGCGTCCGCGGTGCGCGGTGCGCGGTGCGGTG 301  
Db 63153 GCAGTGGCGCGCGCGCGCGCGGTGACGAGTGGCGGGGACCGCGCGCGCGCGCGCG 63212  
QY 302 CGCGGGTGTAGCG 361  
Db 63213 CGGTCCCGCGGTGACGAGGAGCGTGCCTCGGGTCCAGGACGGTCCGCGCGCGCGCG 63272  
QY 362 GACCGATCCGCGAGAGTTCTGTCAGGTGCGCGCTCTTCGCGCTCCGCGCGGTACGCGACA 421  
Db 63273 GCGCGCGCGCGCGCAACCTCGCGCGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 63326  
QY 422 CCCGCTTCCGCTTTCAGCGAGTTTCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 481  
Db 63327 GCGGCTCCGCTCCG 63386  
QY 482 CGACGTCGCGG 492  
Db 63387 CCACGAGGACG 63397

RESULT 33  
AAC65299/c  
ID AAC65299 standard; CDNA; 3510 BP.  
XX  
AC AAC65299;  
XX  
DT 09-FEB-2001 (first entry)  
XX  
DE Maize Scarecrow nucleotide sequence.  
XX  
KW Scarecrow gene; SCR gene; plant; regulatory gene; SCARECROW-like; SCL;  
XX transgenic plant; cell division; molecular marker; herbicide resistance;  
XX salt resistance; pathogen resistance; insect resistance; ss.  
XX  
OS Zea mays.  
XX  
PN WO200053723-A2.  
XX  
PD 14-SEP-2000.  
XX  
PF 07-MAR-2000; 2000WO-US005875.  
XX  
PR 10-MAR-1999; 99US-00265585.  
XX  
PA (UUNY ) UNIV NEW YORK STATE.  
XX  
PI Benfey PN, Di Laurenzio L, Wysocka-Diller J, Malamy JE, Pysh L;  
PI Helariutta Y, Bruce W, Lim J;  
XX WPI; 2000-594315/56.  
XX  
DR P-PSDB; AAB28595.  
XX  
XX Scarecrow gene useful for producing transgenic plants expressing genes  
XX whose product increases starch, lignin or cellulose biosynthesis and  
XX confers herbicide, pathogen or insect resistance.  
XX  
XX Disclosure; Fig 25; 200pp; English.  
XX The present sequence is given in a specification relating to the  
XX structure and function of a regulatory gene SCARECROW (SCR). SCARECROW-  
XX like (SCL) genes encoding proteins containing an amino acid sequence  
XX similar to the sequence of MOTIF III (VHII) of Arabidopsis SCR protein  
XX are disclosed. SCR and SCL genes are useful for producing transgenic  
XX plants whose cell division is modified and root and/or stem development  
XX and gravitropism of stem or hypocotyl is altered. Cell division is  
XX increased or decreased in roots resulting in thicker or thinner root  
XX development. The transgenic plants are useful for expressing a gene of  
XX interest encoding a gene product that confers herbicide, salt, pathogen  
XX or insect resistance in root or embryos and genes encoding starch, lignin  
XX or cellulose biosynthesis in shoots. The SCR gene also confers less  
XX susceptibility to lodging in the transgenic plants than a wild-type  
XX plant. SCR gene sequences are also useful as molecular markers for a  
XX quantitative trait e.g. root or gravitropism trait in molecular breeding  
XX of crop plants  
SQ Sequence 3510 BP; 651 A; 1225 C; 946 G; 688 T; 0 U; 0 Other;  
Query Match 10.9%; Score 54.4; DB 3; Length 3510;  
Best Local Similarity 49.0%; Pred. No. 0.23;  
Matches 145; Conservative 0; Mismatches 151; Indels 0; Gaps 0;  
QY 200 CTTTCGGCATGATCTGTCGGCGTTCAGCGAGAGAGCGCGGTATCTCCACCGCGT 259  
Db 3111 CCAGCGGTGAGCAGCAGAGGTCTCTCCACCGAGCTTCAGCGCGCGCTTCTCTCCA 3052  
QY 260 CGGCGGTGAACGCGTCCGGGGTCCGACCGGGATGTCGTGCGGGGGTGAAGCGCGCCCT 319  
Db 3051 CGAGCGTGTACCGTCCGAGGGGAAATCGCGAGCAGGAGCGCTGCGCGCGCGCG 2992  
QY 320 GCTTGGCGCGCTGCTGTGTCACACCCAGAGACAGGTCCGAGACCGATGCCGACAGT 379  
Db 2991 TGCCGCGAGCGAGCGCGCGCGGACCGGACTGCGGACGCTTCTCGCGCGAGTCCCGA 2932

QY 380 TCGTCACGGTGGGCTCTTCGCGCTGCGCGGTACGCCACACCCCGCTTGCCTCGGCT 439  
Db 2931 ACTTACGTCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2872  
QY 440 TGAGCGAGTTCAGCAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 495  
Db 2871 ACAGCAGCTGCTGCTCCACGAGTGGCGCTGCGCGCGCTGCTGCTGCGCGG 2816

RESULT 34  
ADA70431  
ID ADA70431 standard; DNA; 2964 BP.  
XX AC  
XX ADA70431;  
DT 20-NOV-2003 (first entry)  
XX DE Rice gene, SEQ ID 3754.  
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
XX KW gene; ds.  
XX OS Oryza sativa.  
XX FN WO2003000898-A1.  
XX PD  
XX PF 03-JAN-2003.  
XX PF 22-JUN-2001; 2001WO-IB001105.  
XX PR 22-JUN-2001; 2001WO-IB001105.  
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tac Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175230/17.  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 6; SEQ ID NO 3754; 899pp; English.  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.  
XX Sequence 2964 BP; 671 A; 726 C; 926 G; 636 T; 0 U; 5 Other;  
Query Match 10.7%; Score 53.6; DB 7; Length 2964;  
Best Local Similarity 45.7%; Pred. No. 0.33;  
Matches 170; Conservative 4; Mismatches 198; Indels 0; Gaps 0;

QY 129 GGCTCAGCTCGTCACGCTGACGCGGGGACGTACAGATCCACTGCCCGACCGCGG 188  
Db 120 GGCCAGCGCGCTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 179  
QY 189 GAACTCTGCTCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTA 248  
Db 180 GCTTGTGACAAGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATCA 239  
QY 249 GTCCACCGCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGATG 308

Db 240 GCAGAGTACCTGSGCTGSGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 299  
QY 309 GAGCGCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 368  
Db 300 GAGCG 359  
QY 369 GCGCGAGAGTTCGTACGCGTGGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCTT 428  
Db 360 GCGCGAGAGCTCAA 419  
QY 429 GCGCGCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATGCG 488  
Db 420 GAGCGTGGACACGATG 479  
QY 489 GCGCGCGGAACT 500  
Db 480 CAGCGCGCTGCT 491

RESULT 35  
ACA37987/c  
ID ACA37987 standard; DNA; 1551 BP.  
XX AC  
XX ACA37987;  
DT 19-JUN-2003 (first entry)  
XX DE Prokaryotic essential gene #19644.  
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX KW drug design; gene.  
XX OS Mycobacterium avium.  
XX FN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-023925/02.  
DR P-PSDB; ABU34117.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 25857; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for cell  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)





OS	Homo sapiens.
XX	
XX	
Key	Location/Qualifiers
FF	363..2435
CDS	/*tag=a
FT	/product= "Human ANIC-BP-1 ligand"
FT	
FT	
FT	
XX	
PN	WO200170771-A2.
XX	
PD	27-SEP-2001.
XX	
PP	20-MAR-2001; 2001WO-EP003149.
XX	
PR	21-MAR-2000; 2000EP-00106110.
XX	
PA	(WERE ) MERCK PATENT GMBH.
XX	
PI	Den Daas I, Duecker K, Hock B;
XX	
PI	WPI: 2001-607519/69.
DR	P-PSDB; AAE10857.
XX	
PT	Novel acute neuronal induced calcium binding protein type 1 ligand polypeptides, useful in the treatment of stroke, head trauma, multiple sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord injury.
PT	
PT	
XX	
PS	Claim 4; Page 36-39; 46pp; English.
XX	
CC	The invention relates to human acute neuronal induced calcium binding protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.
CC	Sequences of the invention are useful for treating human diseases
CC	including stroke, head trauma, multiple sclerosis, Parkinson's disease,
CC	Alzheimer's disease and spinal cord injury. They are also useful as
CC	vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound
CC	soluble receptors. Polynucleotides of the invention are useful as
CC	diagnostic reagents, for chromosome localization studies, and as valuable
CC	tools for tissue expression studies. They are also useful in gene
CC	therapy. The present sequence is human ANIC-BP-1 ligand cDNA
XX	
SQ	Sequence 2700 BP; 436 A; 901 C; 857 G; 506 T; 0 U; 0 Other;
	Query Match            10.6%; Score 53.2; DB 4; Length 2700;
	Best Local Similarity 49.6%; Pred. No. 0.39;
	Matches 136; Conservative 0; Mismatches 138; Indels 0; Gaps 0
Qy	218 CGGCGTGTTCCAGCGAAGACGCCGGTAGTCCACCGGTGGGCGGTGAACGCCTCGC 277
Db	326 CGCGGGGGCGCGAGCGCTGCGGTGCTGGGCCAGATGGGGCGGTGGCACGGCCCA 385
Qy	278 GGGTGGCGCACCCGGGATGTGCGTGGCGGGGTGAGCCGGGCCCTCTTTGGCGGTCGTGT 337
Db	386 TGTTGGGTGCGGGTGGCCGGGCCCGCGAAAGTGTCCGCGGGGCGAGCACCGAGGTGC 445
Qy	338 CGCACACCAGAGAACAGGTCCGACCGATGCCGAGAAAGTTCGTACGGTGGGCGCTCT 397
Db	446 AGGACGCCCTGGAGCGCGCTGCGGAGCTGCACAGAGGCTTTGTCCGGCTGAAGCAGG 505
Qy	398 TCGCGGTGCGGCGGTACCCACCCCGCTTGGCCCTTCGGGCTTTGAGCGAGTTACAGCAGG 457
Db	506 CGGGCGGCGGGGCGGTGGGCGCGGCTGGGCCGAGGTCTTCCAACCTGGTGGAGGAGG 565
Qy	458 CGAGCAGGTCGTGCGGATGCCCTTCAGCTCGGC 491
Db	566 CTGTGCTGTGGCGCGTGGGCGCGAGGTAGCC 599
RESULT 39	
ACR37556	
ID	ACA37556 standard; DNA; 1200 BP.
XX	
AC	ACA37556;
XX	

Db	6667	ACAGGCGCGCGCGCGCGTCTCGGCGCGCGCGCGCGTACGACGCGTACGTCGTCGT	6608			
Qy	233	CGAAGACGACGCGT---AGTCCACCGCGTCCGGCGGTGAACGCGTCCGGCGTCCGACCG	289			
Db	6607	GCTCGCGCAGGCGCTCCAGGTCAGGCCGACAGAGGTGCGGCGTCCGCCCGGCGAGGT	6548			
Qy	290	GGATGTCGTGCGCGGGGTGAGCGCGCCCTGCTTGTCGCCGCGTCTGTGTGTCACACCGAG	349			
Db	6547	CGAGCACGCGCGCGTTCGAGCACGGGGCGCGCGCGTTCGAGTCCGTTGTCACACCGGCA	6488			
Qy	350	AGACCAAGTCGCGACCGATGCGCAGAGTTGTCACGTTGGGCGCTTTCGCGCGTCGCGC	409			
Db	6487	AGTCCAGGTGACCCACTCGACCGCGAAACAGCGCGCCGAGCGCAGCGCGTGGCG	6428			
Qy	410	CGTACGCCACACCCGCTTGCCCTCGGCTTGAGCGAGTTACAGAGGCGAGCAGTCCG	469			
Db	6427	GGGCGCGCGCGACCGCGAGGCGTCCACCTCCACGACCGCGCGCGCTCGGGTCTGTAGG	6368			
Qy	470	TGCGGATGCGCTCGACGTCGCGG	493			
Db	6367	CGCGCAGCAGTGTCTGTGCGCG	6344			
RESULT 41						
AAL61191/c						
ID	AAL61191 standard; DNA; 2742 BP.					
XX	AC					
XX	AAL61191;					
XX	22-SEP-2003 (first entry)					
XX	Actinosynnema pretiosum transcriptional activator gene #1.					
DE	Maytansinoid; ansamitocin; antitumour; transcriptional activator; gene;					
XX	Gs.					
XX	Actinosynnema pretiosum.					
XX	WO2003045312-A2.					
XX	05-JUN-2003.					
XX	21-NOV-2002; 2002WO-US037547.					
XX	21-NOV-2001; 2001US-0332158P.					
XX	(UNIW ) UNIV WASHINGTON.					
XX	Floss HG, Yu T, Leistner E;					
XX	WPI; 2003-493374/46.					
XX	Novel maytansinoid produced by bacterial host cell transformed with					
PT	expression vector comprising open reading frame from ansamitocin gene					
PT	cluster I of Actinosynnema pretiosum, useful as antitumor agent.					
XX	Disclosure; Page 81-82; 160pp; English.					
XX	The invention relates to maytansinoid produced by bacterial host cell					
CC	transformed with expression vector comprising open reading frame from					
CC	ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is					
CC	useful as a potent antitumor agent. The present sequence is A. pretiosum					
CC	ansamitocin gene cluster I transcriptional activator gene					
XX	Sequence 2742 BP; 288 A; 1096 C; 1054 G; 304 T; 0 U; 0 Other;					
SQ	Query Match 10.3%; Score 51.4; DB 7; Length 2742;					
Best Local Similarity 51.1%; Pred. No. 0.83;						
Matches 121; Conservative 0; Mismatches 116; Indels 0; Gaps 0;						
Qy	257	CGTGGCGGTGAACGCGTCCGGGTGCGCACCGGATGTCGTCGCGCGGTGACCGCGC	316			





CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC on a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 2010 BP; 309 A; 711 C; 718 G; 272 T; 0 U; 0 Other;

Query Match 10.2%; Score 51.2; DB 7; Length 2010;  
Best Local Similarity 53.3%; Pred. No. 0.92;  
Matches 130; Conservative 0; Mismatches 113; Indels 1; Gaps 1;  
QY 24 AACGATCGAGATACCGGAGAGGACACCCCGGCTAGTCGGGTAGACGCTGGCGCG 83  
DB 1612 AACCTGCATACCGACGCTATACGAAACCGCGCGCGCGCGCTGCGCGCGCGCG 1671  
QY 84 AAGCGGTACGCGCTTCGAGGCTCAGCGGCGGCGGACACCGCGCGGCTCAGCTCGTCA 143  
DB 1672 GAGACG-ACCACGTCGGCTTCGAGGCTCGGCTCGGCTCGGCTCGGCTCGGCTC 1730  
QY 144 GTATACCGGGGAGCGTACAGGATCCACTGTCCCGCAGCGCGGCGGAACTCTGCTCTT 203  
DB 1731 GATCGGCGCGGCGGCGGCTTACCGCGCGGCGGCGGCTCGGCTCGGCGGCGGCA 1790  
QY 204 CGCATGATCTCGTGGCGGCTTCAGGCGGAGAGCAGCGCGCTAGTCCACCGCGTGGG 263  
DB 1791 CGTGGCGCGCTCGGCGGCTTACGTTCCGGAATGGGCGACGCTGTTCCAGGTGTCGG 1850  
QY 264 CGTG 267  
DB 1851 CGTG 1854

RESULT 44  
AAT76906/c  
ID AAT76906 standard; DNA; 546 BP.

XX AC AAT76906;

XX DT 02-JUN-1998 (first entry)

XX S. glaucescens acbD DNA fragment.

XX Aarbose biosynthesis; acbA gene; acbB gene; acbC gene; acbD gene;  
XX acbE gene; acbF gene; enzyme; alpha-amylase inhibitor; treatment;  
XX diabetes; ss.

XX Streptomyces glaucescens.

XX DE19622783-A1.

XX 11-DEC-1997.

XX 07-JUN-1996; 96DE-01022783.

PR 07-JUN-1996; 96DE-01022783.

XX (FARH ) HOECHST AG.

XX Decker H;

XX WPI; 1998-033827/04.

XX Recombinant DNA molecule comprising genes for biosynthesis of acarbose -  
XX an alpha-amylase inhibitor useful in treatment of diabetes.

XX Example 2; Page 7; 35pp; German.

XX This sequence encodes a fragment of the Streptomyces glaucescens GLA.0  
XX acbD gene which has been generated via PCR. The acbD gene is involved in  
XX the acarbose biosynthesis pathway. This gene and other acarbose  
XX biosynthesis genes such as those represented in AAT76903 are useful for  
XX producing acarbose, which is an alpha -amylase inhibitor useful in the  
XX treatment of diabetes

XX Sequence 546 BP; 69 A; 235 C; 168 G; 74 T; 0 U; 0 Other;

Query Match 10.1%; Score 50.6; DB 2; Length 546;

Best Local Similarity 49.8%; Pred. No. 1.3;

Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 239 GCACGCGTATGTCACCGCGCTCGGGGTGAAACGCTCCGGGTGCGCACCGGATGTGCG 298

DB 383 GGAGGAGCGCTATACCTCGTGGTGAGACGTGACGAGGGTGGCGACACCGTGGCGGA 324

QY 299 TGCCGCGGCTGAGCGGCGCTGCTTGGCGCGGCTCGTGTGCGCACACCCAGGACCGAGT 358

DB 323 CGCGGCGCTCGAGCAGGACCTGGGTGCCAGCAGCTTGGTGGCGGTGAAGGCGCGCTGT 264

QY 359 CCGGACCGATGCGCGCAGAAAGTTGTCACGTTGGGCTCTTGGCGCTCGCGCGTACGCCA 418

DB 263 CGGTGATGAGCGGTGCGACGTGCGACTCGGCCCGGAAAGTGCAGATGCTGTCGCGCG 204

QY 419 CCACCGCTTCCCTCGGCGCTTACGCGAGTTTCAGCAGGCGGCGAGCAGTCCGTGGGATGC 478

DB 203 CGGCGAGGCTGCGACGAGCGGCTGTCGACACGCTGCCCTGACGAGGTGAGGCGCG 144

QY 479 CCTCGAGCTCGCGCGCG 495

DB 143 GATGTCACGACCGCG 127

RESULT 45

ADA70563/c

ID ADA70563 standard; DNA; 2697 BP.

XX AC ADA70563;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 3886.

XX Plant; bacterial infection; fungal infection; viral infection; rice;

XX gene; ds.

XX Oryza sativa.

XX WO2003000898-A1.

XX PD 03-JAN-2003.

XX 22-JUN-2001; 2001WO-IB001105.

XX 22-JUN-2001; 2001WO-IB001105.

XX (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX  
XX  
XX Identifying at least one gene involved in plant resistance or response to  
XX pathogenic infection for conferring resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection by determining or detecting plant  
XX gene expression.  
XX  
XX Claim 6; SEQ ID NO 3886; 899pp; English.  
XX  
XX The present invention relates to a method (M1) for identifying genes  
XX involved in plant resistance or response to pathogenic infection. M1  
XX comprises identifying a gene whose expression is significantly altered in  
XX the incompatible interaction of plant gene expression relative to  
XX expression of the gene in an uninfected plant, in a mutant plant that  
XX does not express a gene associated with response to pathogenic infection,  
XX or in a corresponding incompatible or compatible interaction. (M1) is  
XX useful for conferring resistance to resistance or tolerance to a plant to  
XX bacterial, fungal or viral infection. The present sequence was used to  
XX illustrate the invention.  
XX  
XX Sequence 2697 BP; 446 A; 903 C; 922 G; 426 T; 0 U; 0 Other;  
XX  
XX Query Match 10.1%; Score 50.6; DB 7; Length 2697;  
XX Best Local Similarity 48.0%; Pred. No. 1.2; Indels 6; Gaps 1;  
XX Matches 180; Conservative 0; Mismatches 189; Indels 6; Gaps 1;  
XX  
XX QY 118 CGGACACCGCGGGTCACTGTCACGTCGTACGCGGGGAGCTACAGATCCACTGTCG 177  
XX Db 2528 CGCACCACGGTGGGCATGCTCGGCGGAGCGGAGGCTCTCGTGAAGCGCAGAGCGCG 2469  
XX  
XX QY 178 CCAGCCCGCGGAACCTGCTGCTTCCATGATCTCGTGGGCTGTTCCAGCGCAAG 237  
XX Db 2468 ACGCGGACGGTGGTGGCTTGGCCCTCGTGCAGCGGCCCTCCAGCTCGCTCGACG 2409  
XX  
XX QY 238 AGCAGCGGTAGTCACCGCGTCCGGCGTGAACCGGTCCGGGTCGCGACCGGATGTC 297  
XX Db 2408 AGGTCCAGGTATCTCGCTGCTCATGGAGCTCCAGTCCAGTCCGGAAGTACTCTCG 2349  
XX  
XX QY 298 GTGCGCGGGGTAGCGCGCCCTGCTGCGCGGTCGTGTCGACACCCAGGACGAGG 357  
XX Db 2348 CGCGCGCGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2289  
XX  
XX QY 358 TCGGACCGATGCCGAGAGTTCGTACCGTGGCGCTCTTCGCGGTCGCGCGCTAGCGCC 417  
XX Db 2288 TCCGAGTCTCGCGCTCGCGGACGCGGAGCTGTTGTCGCGGCGC-----CGCC 2235  
XX  
XX QY 418 ACCACCGCTGCTCGCGCTTGGCGGTTGAGCGAGTTACAGCGGCGAGGTCGCTCGGATG 477  
XX Db 2234 GCTCTGCTGCGCGGTTCTTCTCCCGTGTATCAGCTCGAGCAGCAGCAGTCCGGAAG 2175  
XX  
XX QY 478 CCTCGACGTGGCGG 492  
XX Db 2174 CTGTAGAGTGGCG 2160  
XX  
XX  
XX RESULT 46  
XX AAL40781/c  
XX ID AAL40781 standard; DNA; 88421 BP.  
XX  
XX AC AAL40781;  
XX  
XX 03-OCT-2002 (first entry)  
XX  
XX 88421nt genomic DNA of ramoplanin producing Actinoplanes sp.  
XX  
XX Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
XX biosynthesis gene cluster; bioengineering; peptide synthetase module;  
XX adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;  
XX chlorinate; lipopeptide; gene; ds.  
XX  
XX Actinoplanes sp.

XX Key Location/Qualifiers  
XX CDS 2077..3078  
XX /tag= a  
XX /product= "Protein of ORF 1"  
XX 3118..4032  
XX /tag= b  
XX /product= "Protein of ORF 2"  
XX 4038..5048  
XX /tag= c  
XX /product= "Protein of ORF 3"  
XX complement(6665..5814)  
XX /tag= d  
XX /product= "Protein of ORF 4"  
XX complement(7703..6693)  
XX /tag= e  
XX /product= "Protein of ORF 5"  
XX complement(9464..8130)  
XX /tag= f  
XX /product= "Protein of ORF 6"  
XX 9691..10761  
XX /tag= g  
XX /product= "Protein of ORF 7"  
XX complement(12751..10829)  
XX /tag= h  
XX /product= "Protein of ORF 8"  
XX complement(13617..12802)  
XX /tag= i  
XX /product= "Protein of ORF 9"  
XX complement(15203..13614)  
XX /tag= j  
XX /product= "Protein of ORF 10"  
XX 15591..15863  
XX /tag= k  
XX /product= "Protein of ORF 11"  
XX 15880..19035  
XX /tag= l  
XX /product= "Protein of ORF 12"  
XX 19032..39713  
XX /tag= m  
XX /product= "Protein of ORF 13"  
XX 39713..65800  
XX /tag= n  
XX /product= "Protein of ORF 14"  
XX 65826..66530  
XX /tag= o  
XX /product= "Protein of ORF 15"  
XX 66546..67370  
XX /tag= p  
XX /product= "Protein of ORF 16"  
XX 67384..70059  
XX /tag= q  
XX /product= "Protein of ORF 17"  
XX 70099..70662  
XX /tag= r  
XX /product= "Protein of ORF 18"  
XX 70659..71906  
XX /tag= s  
XX /product= "Protein of ORF 19"  
XX complement(73439..71964)  
XX /tag= t  
XX /product= "Protein of ORF 20"  
XX complement(74216..73563)  
XX /tag= u  
XX /product= "Protein of ORF 21"  
XX complement(75424..74213)  
XX /tag= v  
XX /product= "Protein of ORF 22"  
XX 75535..76464  
XX /tag= w  
XX /product= "Protein of ORF 23"  
XX complement(78110..76449)  
XX /tag= x

FT CDS /product= "Protein of ORF 24"  
FT complement(79864..78107)  
FT /\*tag= y  
FT CDS /product= "Protein of ORF 25"  
FT complement(81624..79861)  
FT /\*tag= z  
FT CDS /product= "Protein of ORF 26"  
FT complement(81909..81682)  
FT /\*tag= aa  
FT CDS /product= "Protein of ORF 27"  
FT complement(82346..82062)  
FT /\*tag= ab  
FT CDS /product= "Protein of ORF 28"  
FT 82587..84446  
FT /\*tag= ac  
FT CDS /product= "Protein of ORF 29"  
FT 84481..85548  
FT /\*tag= ad  
FT CDS /product= "Protein of ORF 30"  
FT 85556..86845  
FT /\*tag= ae  
FT CDS /product= "Protein of ORF 31"  
FT 87372..86803  
FT /\*tag= af  
FT CDS /product= "Protein of ORF 32"  
FT 87494..88420  
FT /\*tag= ag  
FT CDS /product= "Protein of ORF 33"  
FT  
XX WO200231155-A2.

18-APR-2002.

15-OCT-2001; 2001WO-CA001462.

13-OCT-2000; 2000US-0239924P.

12-APR-2001; 2001US-0283296P.

24-JUL-2001; 2001US-00910813.

(ECOP-) ECOPIA BIOSCIENCES INC.

Farnet CM, Zazopoulos E, Staffa A;

WPI: 2002-435445/46.

P-PSDB: AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,

RAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,

RAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,

RAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,

RAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.

Novel isolated ramoplanin biosynthetic pathway polypeptide useful for

chemically modifying biological molecule that is a substrate for a

polypeptide encoded by a ramoplanin biosynthesis gene cluster.

Disclosure; Page 87-135; 212pp; English.

The invention relates to an isolated ramoplanin biosynthetic pathway  
polypeptide selected from a polypeptide of open reading frames (ORF) 1-  
32. The isolated polypeptides are useful for chemically modifying a  
biological molecule that is a substrate for a polypeptide encoded by a  
ramoplanin biosynthesis gene cluster, by contacting the biological  
molecule with the isolated polypeptide, where the polypeptide chemically  
modifies the biological molecule. The method comprises contacting the  
biological molecule with at least two different polypeptides encoded by  
ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
gene cluster comprising the ORFs is useful as a substrate for  
bioengineering of antibiotic structures. An isolated polypeptide or its  
encoding nucleic acid sequence is useful for generating derivatives of  
antibiotics, for improving production or for producing variants of other  
antibiotics of the peptide class. The isolated polypeptides are useful  
for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain  
in conjunction with other peptide synthetase modules and allowing the

CC incorporation of Thr into a peptide antibiotic precursor, for modifying  
CC fatty acid structure and/or enhancing fatty acid incorporation into the  
CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
CC ramoplanin or its variants and derivatives, for enhancing uptake of  
CC precursors for ramoplanin biosynthesis, for enhancing production of  
CC a peptide antibiotic or its variants or derivatives, to chlorinate HPG of  
CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
CC a peptide antibiotic precursor, and for designing specific nucleotide  
CC probes and primers for identifying and isolating putative lipopeptide  
CC -producing microorganisms. this polynucleotide sequence represents the  
CC 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.  
XX microorganism of the invention

SQ Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;

Query Match 10.1%; Score 50.6; DB 6; Length 88421;

Best Local Similarity 45.3%; Pred. No. 0.87;

Matches 224; Conservative 0; Mismatches 269; Indels 2; Gaps 1;

Qy 6 CCGACCGTGTCCCGGCAACACGAGTCGAGATACGGGAGAGGAACACCCCGGTAGTCC 65  
Db 71491 CCGAACGTGCCAGGCGACGGGTTTACCGCGGCATGTGCAGGCCGAGCCGTCAGC 71432  
Qy 66 GGGTAGACGTTGGCGCGGAGGGGTACGGCCCTTCACCGGTTCAGCGGCGGGCGGACACC 125  
Db 71431 GCGAACACAGAGCACCGCAACTCGGCCCGCGCGGTGAAACGCCAGAACTCCATC 71372  
Qy 126 GGGGGGTACGTCTGTACGTGTACGGGG--GGACGTACAGGATCCACTGTCCGCCAGCC 183  
Db 71371 CAGCGTCCGGTTGCAGCGGACACCGCGCAGTTTACACAGAGCCAGGCCCGCGGCA 71312  
Qy 184 GGGCGGAATCTCTGCTCTTCGCGCATGATCTGTGGGGTGTTCAGGGGAAGAGCAGC 243  
Db 71311 CCGCGCGCCACGCGGGGAAACCCCGCATCCCGCTGCGGAGGAGCAGCAGCAGC 71252  
Qy 244 GCGTAGTCCACCGTCGCGCGGTGAACCGCTCCGGGTGCGACCGGATGTCCGTGCGCG 303  
Db 71251 GGGCCGAGCAGACACCGCGGACAGCTTGGCCCGCTGCCAGCCGATCAGCAGCG 71192  
Qy 304 GGGGTGAGCCGCGCTGTGTTGGCGGGGTCTGTCTGCGAACHCCAGAGACAGGTCCGGA 363  
Db 71191 GCCAGCAGCGCGGGTCTCCGTTCACGGGAGAGATCGCAGACACCGCGCGGACCG 71132  
Qy 364 CCGATGCCGAGAAAGTTTCGTACGGTGGGCTCTTCGCGCTCGCGCGGTACGCCACCC 423  
Db 71131 AGCAGTCCAGTTCATCCCGTCCGCGCAGATCAGCACCGGTGCGGGCGGACGAGC 71072  
Qy 424 CGTTGCGCTCGGCTTTGACGCGAGTTTACGAGGCGGAGCAGGTGCGGTGCGGATGCC 483  
Db 71071 GCGTCGCGCGCGCGCGGTGCGGCGCAGCAGCAGCCGTCGCGCCACACGCGTCAG 71012  
Qy 484 ACCTCGCGCGGCAAC 498  
Db 71011 ACCCGAGGCGGAGC 70997

RESULT 47

AAQ10190/c

ID AAQ10190 standard; DNA; 23666 BP.

XX AAQ10190;

XX 25-MAR-2003 (revised)

DT 27-MAR-1991 (first entry)

XX Cephalosporin antibiotic biosynthetic genes.

XX cephalosporin; antibiotic; S-(L-alpha-aminoadipyl)-L-cysteiny]-D-;  
KW valine synthetase; isopenicillin N synthetase; isopenicillin N epimerase;  
KW deacetoxycephalosporin C synthetase; beta-lactamase;  
KW deacetoxycephalosporin C hydroxylase; ss.  
XX Lysobacter lactamgenus.  
OS



CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1248 BP; 157 A; 451 C; 474 G; 166 T; 0 U; 0 Other;

Query Match 10.0%; Score 50.2; DB 7; Length 1248;  
Best Local Similarity 50.7%; Pred. No. 1.5;  
Matches 180; Conservative 0; Mismatches 163; Indels 12; Gaps 2;

QY 158 CGTACAGGATCCACTGTCGCGCAGCCGCGGAATCCCTGCTTCGCGCATGATCTG 217  
Db 616 CGCGGACGATCACTCGCGCTCGCGGAGCTACCGAATCGTTGCGCCAGAGCGCG 557  
QY 218 CGCGGTGTTCCAGGCGAAGAGACGCGGTAGTCCACCGCTCGGCGGTGAACGGGTCCG 277  
Db 556 CGCGCGCCACCCACACACAGTGGGACTGTCTGAGCGCGGAGGTATCAGCAGTGT 497  
QY 278 GGTGCGCACCGGAGTGTGCTGCCG-----GGGTAGCGCGCCCTCTTGGCGCGG 331  
Db 496 CGAGTTGGCCACGACACCGCTCGCGGTCCGGCCAGCGCGCGCGCGAGGTCCGCT 437  
QY 332 TCCTGTGCGCACACCCAGGAGACAGGTCCGGACCGATGCCGAGAGTTTCGTACCGGTGG 391  
Db 436 CGATGACGTTGATCTCGCGACCCGCTCAGCTGTGTCGGGACAGGTTGAGCAGCACA 377  
QY 392 CGTCTTTCGCGTCCGCGCTAGCCAC-----CACCGCTTGCCTCGGCTTACGG 445  
Db 376 CGACGCCCGCGCGACGGCGTCCGACAGTGGCGCACGTGGCGCACGTGCTCGACCTCCAGGG 317  
QY 446 AGTTCAGCAGGCGGAGCAGGTCCGTGCGGATGCCCTCGACGTCCGCGCGGAACCT 500  
Db 316 CGGCCAGCGCGCGCGCTTGGCGGCGAGCGCGCGACCGACGCGCGCGTCCAT 262

RESULT 49  
AA02914/c  
ID AAX02914 standard; DNA; 1605 BP.

XX AAX02914;

XX 20-MAY-1999 (first entry)

XX Z. mays cck1 DNA coding region.

XX Cytokinin oxidase; cck1; transgenic plant; altered growth behaviour;  
KW cytokinin-associated pathogenesis; resistance; fungi; nematode; assay;  
KW grain yield; secondary growth; metabolism; senescence; ss.

XX Zea mays.

XX WO9906571-A1.

XX 11-FEB-1999.

XX 30-JUL-1998; 98WO-US015844.  
PF 30-JUL-1997; 97US-0054268P.  
PR 29-JUL-1998; 98US-00124541.  
XX (UMOR ) UNIV MISSOURI.  
XX Morris RO;  
PI WPI; 1999-153800/13.  
DR P-PSDB; AAW93007.  
XX New cytokinin oxidase from maize - used to generate transgenic plants  
FT with, e.g. better disease resistance and growth characteristics.  
XX Claim 7b; Page 64-69; 140pp; English.  
PS This sequence encodes a Zea mays cytokinin oxidase, cck1. This gene is  
XX used to generate transgenic plants in which cytokinin-associated  
CC pathogenesis or growth behaviour is altered. Particularly applications  
CC include generation of plants with increased resistance to fungi and  
CC nematodes, increased grain yield and superior secondary growth  
CC properties. Host cells are used for production of recombinant cck1  
CC nucleic acid which is useful in an assay for determining cytokinin  
CC concentrations, and for its studying effects on plant growth and  
CC metabolism, including senescence  
XX Sequence 1605 BP; 236 A; 560 C; 561 G; 248 T; 0 U; 0 Other;

Query Match 10.0%; Score 50.2; DB 2; Length 1605;  
Best Local Similarity 51.0%; Pred. No. 1.4;  
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCACTGTCCGCGCAGCCGCGGAACTCTCTCTCTTCCGCGCATGATCTCGTGGGTG 224  
Db 1194 GAGCCACGGGTGCGGACCCCGCCACAGCCCGAGCTTGTGAGCGCCACCTCTCGCGGTG 1135  
QY 225 GTTCCAG3CGAAGAGCAGCGTAGTCCACCGGTGCGGCGGTGAACGCGTCCGCGGTGCG 284  
Db 1134 CACCGGTTCAGGAAACCGCGGTAGGCA---CGTGGGTGAGACGCGAAC---CCCTC 1081  
QY 285 CACCGGATGTGCTGCGCGGGGTGAGCCGCGCGCTTGTGGCGCGCGCTGTGTGCGCAC 344  
Db 1080 CACGTAGTTCAGCGTGCACGACGCGGAGCTCTGTGTCACCGTGGTGGCGCGC 1021  
QY 345 CCAGGAGACAGGTCCGACCGGATGCCGAGAGTTCTGTCACGTGGCGGTCTTCGCGCT 404  
Db 1020 CGTGGCGTGTGTGAGTGTGAGCGTGGCTGTGATCTGTACCGTGGTGGGTTCGCTC 961  
QY 405 CGCGCGGTACCGCCACCGCGCTTCCCTCGGCTTGAAGGAGTTACAGAGGCGAGCAG 464  
Db 960 CCGCGGAGCGGACGATCCGCGGCGACGTGGCGGTGCGTGAAGAACCCCGTGTTCGCGCAG 901  
QY 465 GTCGTTGCGGATGCCCTCGACGTCCGCGCGGACCC 499  
Db 900 GTCGTTGCGGATGCCCTCGACGTCCGCGCGGACCC 866

RESULT 50

AA02913/c  
ID AAX02913 standard; DNA; 6733 BP.

XX AAX02913;

XX 20-MAY-1999 (first entry)

XX Z. mays cck1 DNA.

XX Cytokinin oxidase; cck1; transgenic plant; altered growth behaviour;  
KW cytokinin-associated pathogenesis; resistance; fungi; nematode; assay;  
KW grain yield; secondary growth; metabolism; senescence; ss.

2928 CGTGGCGTTGCTAGTTGAGCGTGGCCCTCGATGCTACACGGTGGTGGGTTCCGGCTC 2869  
405 CGCGCGGTACGCCACCCCGCTTCCCTCGCCCTTGAGCGAGTTCAGCAGGCGGAGCAG 464  
2868 CCGCGGAGCGCGACGATCGCGGCGACGTCGCGCGGACGTCGCGTGAAGAACCCCGCTTCCGGCAG 2809  
465 GTCGGTGGGATGCCCTCGACGTCGGCGGCGAACC 499  
2808 GTGCGTCCGAGGCTCTGTTACGAAACACCGACC 2774

Search completed: June 27, 2004, 18:18:33  
Job time : 332.98 secs

Zea mays.  
Key Location/Qualifiers  
CDS 1497..3607  
/\*tag= a  
/\*product= "ckx1"  
exon 1497..2111  
/\*tag= b  
/\*number= 1  
intron 2112..2523  
/\*tag= c  
/\*number= 1  
exon 2524..3216  
/\*tag= d  
/\*number= 2  
intron 3217..3310  
/\*tag= e  
/\*number= 2  
exon 3311..3607  
/\*tag= f  
/\*number= 3

W09906571-A1.

11-FEB-1999.

30-JUL-1998; 98WO-US015844.

30-JUL-1997; 97US-0054268P.

29-JUL-1998; 98US-00124541.

(UMOR ) UNIV MISSOURI.

Morris RO;

WPI; 1999-153800/13.

P-PSDB; AAW93007.

New cytokinin oxidase from maize - used to generate transgenic plants with, e.g. better disease resistance and growth characteristics.

Claim 7a; Page 54-63; 140pp; English.

This sequence encodes a Zea mays cytokinin oxidase, ckx1. This gene is used to generate transgenic plants in which cytokinin-associated pathogenesis or growth behaviour is altered. Particularly applications include generation of plants with increased resistance to fungi and nematodes, increased grain yield and superior secondary growth properties. Host cells are used for production of recombinant ckx1 nucleic acid which is useful in an assay for determining cytokinin concentrations, and for its studying effects on plant growth and metabolism, including senescence

Sequence 6733 BP; 1786 A; 1585 C; 1592 G; 1765 T; 0 U; 5 Other;

Query Match 10.0%; Score 50.2; DB 2; Length 6733;  
Best Local Similarity 51.0%; Pred. No. 1.3; Indels 6; Gaps 2;  
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;  
QY 165 GATCCACTGTCGCCAGCCGCGGAACTCTCTCTCGCCATGATCTCGCGCGTG 224  
DB 3102 GAGCCAGGGTGGCGACCCGCGACAGCCCGCAGCTTTGAGCGCCACCTCTCGCGTG 3043  
QY 225 GTTCCAGGGGAGAGCAGCGCGTAGTCACCGCGTCGGCGGTGAACGCGTCGGGGTGG 284  
DB 3042 CACCCGGTCAAGGAGACCGCGTAGGCCA---CGTCGCGCTGGAGCGGAACCC---CTC 2989  
QY 285 CACCGGATGTGGTGGCGGGGTGAGCGCGCCCTCTTGGCCGGGTGCTGTGCGCACAC 344  
DB 2988 CAGTAGCTCAGGTGCGCCAGCAGCAGCGAGCTCTGCTCACCGCGCGCGCGCGC 2929  
QY 345 CCAGGAGACCAAGTCCGGACCGGATGCGCAGAAAGTTGCTACGCTGGCGCTCTTCCCGT 404

SQTGLVRSRTSLAATPPAAKSCSRVTAASEGGGAAAGAAAGAPVSVAGSPLS  
PGVAPVRSHTLSGGCGGKSVALLPAGALQHSRSMMPVAHSPPAATSPGSL  
SSSDGSGSPYPPPHPLPLPHLHGQRPSSGASAGSDPDGFMGLDVGSP  
GDLRAFCSHRNTPEIAETPPADGGGGEGFYCYMTMDRLSHCHRSYRVSGDAQ  
DLRLGRKRTSLTTPARQVPVPPSSASLDEYTLMAATPSGAGRLCPGCPASSXV  
AYHPYEDYDIEIGSHRSSNUGADDGTMPTFGAALAGSGSGCKDDIMPSFA  
SVAPKQTLQPRAAAATAAAPSAGPAGPAPTSAAAGTFFPASGGYKASPAESSED  
SGYRMWCGSLKSEHADKLLPNGLDNLVSPDAVTGTPDFSAALHGGEPJRG  
VPGCCYSLPSYKAPYTCGGSDQVYLMSPVGRILEERLEPQATPGSOAASAF  
AGTQPPHVPVSVPRSGRPEGLGQGRVAPTRLISLEGLPSLMSHEYLPLPFP  
KSPGEYINDFEGECARLSPPALLASASSSLSSASSPASSLSGTPTGSSDSRQ  
RPSLDYNNLDGEPKSPKPGARSHGPHVSLDGLLSPASSPVPLPSPASSSL  
QPPPPPPAGLYRLPAPASAVATAGGGAASSLSDGNDGDTYEMAFVAAATPQPI  
AAPKPEAAVASPTTSGVRLSMEQVSGVEAFLOASQPPDPRHGAIVRADPQGRK  
RHSSETFTTTPVSPFAHNRHNSASVENYLRKSEGGVGVGGGDEPPTS  
PQLOPAPPLAQPRPMTGQGLVCGSGSPMRERTSAGFQNGLYAIADVRE  
PGLPPOPPPLPQPGDKSWGRTRSLGLISAVGVSTGGCGGCGPCCALPPANT  
VASIDFLSHLKEATIVK  
223..225  
variation  
/gene="IRS2"  
/note="compared to GenBank Accession Number AF322114"  
/replace=""

ORIGIN  
Query Match 10.5%; Score 54.8; DB 9; Length 4286;  
Best Local Similarity 46.2%; Pred. No. 41;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCCACTCGCGGGGTGACCAAGTTCAGACAGTGGAGTTGAGCAGCGCTCATCG 105  
DB 1037 AGTTCGGCGCGCGAGTAAGACCAATCGTCGGGTCGTGCGGCCACGCCCATCAGCG 1096  
QY 106 CGCGCTGGCGCGATCGCGAACCACCGGCCAGGATCAGTGCAGCAGCGCGATGCGCG 165  
DB 1097 TCCCGCGCGCGCGCGCCACCAACCACTGTCACCTGCCCGCCACGACGCGCGCTGG 1156  
QY 166 CCGTGCATCTGACGCTGCGCGCGAGTTCGATCTGTCGCGCGGCTCCACACACCGAGGAA 225  
DB 1157 TCGCGCGCTGCGCGACCGACGCTGCGCGCCACCGCGCGCGCGCGCAAGTGCAGCTCGT 1216  
QY 226 ACTGCTCGTGGCAGCGCGCGCGCGCGCGCGAGTCTGTCACAGGCTGGGTCACGGA 285  
DB 1217 GCGCGGTGCGACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1276  
QY 286 CCGTCTCGCGGTCGGGAGCGGAAACGCGCAGCGCTCAGATCCCTCTCAGTCCGATCGGCT 345  
DB 1277 GCGCCAGCGCGTGTGCTGGTGGAGC-CCCTGAGCGCGCGCGCGCGCGCGCGCGCG 1335  
QY 346 CAGTCCCGCTGTCGCCCTTGGCTGGAGGATAGCGGTTACAGCAGCGCGCACCAAG 405  
DB 1336 CTGAGCGCTGCGCACACCTGAGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCTG 1395  
QY 406 CGCGCGCGCGCGCGCGCGGTCAGCGATCCGCTCGATGACCGCGGCTGCGGGTTCGGG 465  
DB 1396 CTGCGCGCGCGCGCGCGCGCTGCAACACAGCGCTCCTATGTCATGCCCTGCGCGCTG 1455  
QY 466 GCGGTGCGGAGATCGTACCGCGCGGACCGCTGCGCGCGCGCGCGCGCGCGCGCGG 515  
DB 1456 CGCGCGCGCGCGCGCGCGCGCTCCCTGTCGTCGTCGCGCGCGCGCGCGCGCGCGG 1505

RESULT 24  
AF385932 4289 bp DNA linear PRI 25-JUN-2001  
LOCUS Homo sapiens insulin receptor substrate 2 insertion mutant (IRS2)  
DEFINITION gene, partial cds.  
ACCESSION AF385932  
VERSION AF385932.1 GI:14537853  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 4289)  
AUTHORS Heyne, B.  
TITLE Two insertions in insulin receptor substrate 2 (N28\_H29insN);  
N28\_H29insNN)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4289)  
AUTHORS Heyne, B., Gehrisch, S. and Jaross, W.  
TITLE Direct Submission  
JOURNAL Submitted (25-MAY-2001) Institut fuer Klinische Chemie und  
Laboratoriumsmedizin, Universitaetsklinikum der Technischen  
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany  
FEATURES  
Location/Qualifiers  
1..4289  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/map="13q34"  
/note="isolated from a patient with type 2 diabetes  
mellitus"  
139..>4156  
/gene="IRS2"  
/product="insulin receptor substrate 2 insertion mutant"  
139..>4156  
/gene="IRS2"  
/codon\_start=1  
/product="insulin receptor substrate 2 insertion mutant"  
/protein\_id="AAK66750.1"  
/db\_xref="GI:14537854"  
/translation="MASPPRHGPPGASGDGPNLNNNNNNNNNNHVRKCYLRKQKH  
GKRRFVLPAGAGDEATAGGSAQPPRLVEYSEKKWRKAGAPKRVIALDCCIN  
INKRADAKHYKLIATYKDEYFAVAENEOEGWYRALTDLVSEGAAGADAPAAA  
PAASCASILPAGLGSAGAGAEADSYLVAPATAAYEVQVNLKPKGQSKNLTVG  
YRUCLSARTIGFVLCNCPQSVTLQLMNIRGSHSDFFIEVGRSAVTGPELWMAQ  
DDVSAQNTHTILEAMKALKEFEPKRSKQSSGSSATHFVSGARRHHVNLVLP  
PSQTLVRSRTSLAATPPAAKSCSRVTAASEGGGAAAGAAAGAPVSVAGSPL  
SPGVAPVRSHTLSGGCGGKSVALLPAGALQHSRSMMPVAHSPPAATSPGSL  
SSSDGSGSPYPPPHPLPLPHLHGQRPSSGASAGSDPDGFMGLDVGSP  
GDLRAFCSHRNTPEIAETPPADGGGGEGFYCYMTMDRLSHCHRSYRVSGDAQ  
QLDLGRKRTSLTTPARQVPVPPSSASLDEYTLMAATPSGAGRLCPGCPASSXV  
VATHPYEDYDIEIGSHRSSNUGADDGTMPTFGAALAGSGSGCKDDIMPSFA  
SVAPKQTLQPRAAAATAAAPSAGPAGPAPTSAAAGTFFPASGGYKASPAESSED  
SGYRMWCGSLKSEHADKLLPNGLDNLVSPDAVTGTPDFSAALHGGEPJRG  
VPGCCYSLPSYKAPYTCGGSDQVYLMSPVGRILEERLEPQATPGSOAASAF  
AGTQPPHVPVSVPRSGRPEGLGQGRVAPTRLISLEGLPSLMSHEYLPLPFP  
KSPGEYINDFEGECARLSPPALLASASSSLSSASSPASSLSGTPTGSSDSRQ  
RPSLDYNNLDGEPKSPKPGARSHGPHVSLDGLLSPASSPVPLPSPASSSL  
QPPPPPPAGLYRLPAPASAVATAGGGAASSLSDGNDGDTYEMAFVAAATPQPI  
AAPKPEAAVASPTTSGVRLSMEQVSGVEAFLOASQPPDPRHGAIVRADPQGRK  
RHSSETFTTTPVSPFAHNRHNSASVENYLRKSEGGVGVGGGDEPPTS  
PQLOPAPPLAQPRPMTGQGLVCGSGSPMRERTSAGFQNGLYAIADVRE  
PGLPPOPPPLPQPGDKSWGRTRSLGLISAVGVSTGGCGGCGPCCALPPAN  
TVASIDFLSHLKEATIVK"  
223..228  
/gene="IRS2"  
/note="compared to GenBank Accession Number AF322114"  
/replace=""

variation

ORIGIN  
Query Match 10.5%; Score 54.8; DB 9; Length 4289;  
Best Local Similarity 46.2%; Pred. No. 41;  
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;  
QY 46 ACTGCCACTCGCGGGGTGACCAAGTTCAGACAGTGGAGTTGAGCAGCGCTCATCG 105  
DB 1040 AGTTCGGCGCGCGAGTAAGACCAATCGTCGGGTCGTGCGGCCACGCCCATCAGCG 1099  
QY 106 CGCGCTGGCGCGATCGCGAACCACCGGCCAGGATCAGTGCAGCAGCGCGATGCGCG 165  
DB 1100 TCCCGCGCGCGCGCGCGCGCTCCCTGTCACCTGTCGTCGTCGCGCGCGCGCGCTGG 1159

```
QY 166 CCTCGATCTCGAGCGGTGGCGGCGAGCTCGATCTCGTGGCCGCGCTCCACACAGGGGAA 225
Db 1160 TGGCCCGCTCGGACCGAGCGCTGCGCCGCCACCCCGCGCGCGGCGCAAGTCAGTCGT 1219
QY 226 ACTGCTCGGTGGCGAGCGGCGAGCCGCCAGCGGACAGCTCGTCCAGGCTGGCGTCAAGGA 285
Db 1220 GCGGGTGGCGACCGCGAGCGAGCGGCGAGCGGCGGCGGCGGAGCGCGCGCGCG 1279
QY 286 CCTTCGCGGGTGGGAGCGGGAACGCGACGCGTCAAGTCCCTGTCAGTCCGATCGGCT 345
Db 1280 GCGCAGCGCGGTGTCGTGGCTGGGAGC-CCCTGAGCCCCGGCGCGGCGCGCGCC 1338
QY 346 CAGTGGCGGTGCTCCCTTGGCTGGGAGGATAGCGGTTTCAAGCAGCGGACCAACGG 405
Db 1339 CTGAGCGCTCGCACACCTGAGCGCGGTGCGGCGCGCGCGGAGCAAGGTGGCGCTG 1398
QY 406 CGGCGGGGGCGGGGGGTTTCAAGCATCGCTCGATGACCGAGCGGCTCGCGGGTGGG 465
Db 1399 CTGCGCGAGGGGCGCGCTGCAACACAGCGCTTCCATGTCATGCGCGGCGGCGACTCG 1458
QY 466 CGGTGCGGAGATCCGTACCGCGCGGACCGCTCGGCGCGCGCGCGCGCGG 515
Db 1459 CCGCGCGCGCACCGCGCGCGCTCTCTGCTCCAGCAGCGCGCGCACGG 1508
```

```
RESULT 25
AL162497/c
LOCUS
DEFINITION Human DNA sequence from clone Rp11-313L9 on chromosome 13, complete
sequence.
ACCESSION AL162497.20 GI:14329908
VERSION AL162497.20
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 143409)
Direct Submission
Submitted (05-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 8, 2001 this sequence version replaced gi:14280409.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations.
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
Rp11-313L9 is from the library RPI1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-313L9 It may be shorter because we sequence overlapping
```

sections only once, except for a 100 base overlap.  
The true right end of clone Rp11-313L9 is at 143409 in this  
sequence. The true left end of clone Rp11-40B6 is at 100074 in this  
sequence. The true right end of clone Rp11-358F13 is at 100 in this  
sequence.

```
FEATURES
source
1..143409
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/clone="Rp11-313L9"
/clone_lib="RPI1-11.2"
1..350
/note="L1M9A repeat: matches 5421..5789 of consensus"
351..654
/note="L1P81 repeat: matches 5846..6155 of consensus"
655..957
/note="L1M9A repeat: matches 5789..6085 of consensus"
2063..2106
/note="23 copies 2 mer tg 91% conserved"
2063..2106
/note="11 copies 4 mer ttgtg 93% conserved"
2138..2422
/note="L1M3 repeat: matches 6691..7023 of consensus"
2422..3051
/note="L1M4 repeat: matches 7168..7837 of consensus"
3674..4545
/note="L1P4 repeat: matches 5272..6144 of consensus"
5701..5910
/note="7 copies 30 mer 89% conserved"
6490..6613
/note="MIR repeat: matches 92..225 of consensus"
9599..9648
/note="25 copies 2 mer aa 94% conserved"
10251..10559
/note="AluX repeat: matches 2..312 of consensus"
11134..11521
/note="L1M7 repeat: matches 5762..6161 of consensus"
12667..12710
/note="11 copies 4 mer ttgtg 81% conserved"
13210..13309
/note="L2 repeat: matches 2559..2674 of consensus"
13794..14148
/note="THEIC repeat: matches 1..371 of consensus"
14816..15030
/note="MIR repeat: matches 49..256 of consensus"
14982..15037
/note="L2 repeat: matches 2651..2705 of consensus"
15476..15940
/note="L2 repeat: matches 1314..1835 of consensus"
16305..16609
/note="AluX repeat: matches 1..300 of consensus"
16683..16722
/note="20 copies 2 mer tt 77% conserved"
16948..17073
/note="MLTIC repeat: matches 348..464 of consensus"
17074..17360
/note="AluJ repeat: matches 13..299 of consensus"
17361..17661
/note="MLTIC repeat: matches 7..348 of consensus"
17974..18347
/note="L1M1 repeat: matches 5659..6029 of consensus"
18399..18554
/note="3 copies 52 mer 75% conserved"
18400..18549
/note="5 copies 30 mer 73% conserved"
18413..18552
/note="5 copies 28 mer 74% conserved"
18565..18738
/note="L1M2 repeat: matches 6150..6335 of consensus"
19070..19105
/note="9 copies 4 mer gata 88% conserved"
```







COMMENT On or before Oct 26, 2002 this sequence version replaced  
 gi:20520847, gi:20520774, gi:20520824, gi:20520691, gi:20520851,  
 gi:20520775, gi:20520830, gi:20520831, gi:20520832, gi:20520833,  
 gi:20520835, gi:20520841.

FEATURES  
 Location/Qualifiers  
 1..30800  
 /organism="Streptomyces coelicolor A3(2)"  
 /mol\_type="genomic DNA"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 84..89  
 /gene="84\_89"  
 98..973  
 /gene="SCO2256"  
 /note="synonym: panB"  
 98..973  
 /gene="SCO2256"  
 /note="SCC75A.02, panB, 3-methyl-2-oxobutanate  
 hydroxymethyltransferase (EC 2.1.2.11), len: 291 aa.  
 Highly similar to many including: Escherichia coli  
 SW:PANB\_ECOLI(EMBL:L17086) 3-methyl-2-oxobutanate  
 hydroxymethyltransferase (EC 2.1.2.11) (264 aa), fasta  
 scores opt: 711 z-score: 839.4 E(): 0.44.3% identity in  
 262 aa overlap and Mycobacterium tuberculosis  
 SW:PANB\_MYCTU(EMBL:Z70692) 3-methyl-2-oxobutanate  
 hydroxymethyltransferase (EC 2.1.2.11) (281 aa), fasta  
 scores opt: 1019 z-score: 1199.2 E(): 0.59.1% identity in  
 269 aa overlap."  
 /codon\_start=1  
 /transl\_table=11  
 /product="3-methyl-2-oxobutanate hydroxymethyltransferase  
 (EC 2.1.2.11)."  
 /protein\_id="CAB61702.1"  
 /db\_xref="GI:6469241"  
 /db\_xref="GOA:Q9RKS2"  
 /db\_xref="SWISS-PROT:Q9RKS2"  
 /translation="MTQLSRAQTPQPKPADGNRALYGGKINRITVDITAAKRGEXK  
 APMLTAYDAMTASVDESIGIEPMLVGDSSAGNCHGYETVPTIDEMTLSAAVVRGT  
 SRALIVGLDFPGSQEGVEQALSRATLVKEAGVAVLKEGERSHQIELLVESGIP  
 VMAHIGLTPQSVNSGYRQGRBEAAQQLLRDAKAVQDAGAFVLELPAELAAEV  
 TRTLHIPVIGIGAGPDTDAVLVMTDLGLTGRMPKFVKQYDLRKVMGDAAKAYAE  
 DVVGSTFPADSESHYH"  
 1093..2190  
 /gene="SCO2257"  
 /note="synonym: SCC75A.03"  
 1093..2190  
 /gene="SCO2257"  
 /note="SCC75A.03, probable ABC transporter, ATP-binding  
 component, len: 365 aa. Highly similar to many ABC  
 transport proteins e.g. Streptomyces peucetius  
 SW:DRRB\_STRPE(EMBL:M73758) daunorubicin resistance  
 ATP-binding protein (330 aa), fasta scores opt: 922  
 z-score: 1009.0 E(): 0.48.9% identity in 329 aa overlap and  
 Streptomyces antibioticus TR:Q53716(EMBL:L06249)  
 ATP-binding protein OLEC-ORF4 involved in  
 oleandomycin-resistance (325 aa), fasta scores opt: 1883  
 z-score: 2452.8 E(): 0.89.7% identity in 329 aa overlap.  
 Also similar to several Streptomyces coelicolor putative  
 ABC transporter ATP-binding components e.g.  
 TR:CAB42771(EMBL:AL049841) SCE9.25C (315 aa), fasta scores  
 opt: 1031 z-score: 1127.7 E(): 0.54.4% identity in 318 aa  
 overlap. Contains Prosite hits to PS00017 ATP/GTP-binding  
 site motif A (P-loop) and PS00211 ABC transporters family  
 signature. Also contains a Pfam match to entry PF00005  
 ABC\_tran, ABC transporter."  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable ABC transporter, ATP-binding  
 component."  
 /protein\_id="CAB61703.1"  
 /db\_xref="GI:6469242"  
 /db\_xref="GOA:Q9RKS1"  
 /db\_xref="SPTREMBL:Q9RKS1"  
 /translation="MAPAGCRWDVGALSVPPAIVGMKRIIDIPRAADSAVTVRGLVK

HYGETKALGDVLDVRECTVMGVLPNGAGKTLVRLITLTIPDSGOAHVAGYDVVR  
 QPRQRRVIGITQTASVDEKLPQWENLYMIGRLIDSRDARTRADLLRPSLTERA  
 KRPASTYSGEARRLIDLAASMI GHPAFLFDEPTGLDPRTRNEVDEVKRMVGDGV  
 TVLITQYMEAGELASLTVDKRGVIANGIIEELKAKVGGRTRIRPADPLQRPL  
 AAYLDELGITGLASTVDTERTGTVLPILSDEQLTAVVGVAVTARGITVSAVTTLP  
 DEVLSLTGHEASAPQDPAPVTDTRREAVV"

misc\_feature  
 1276..1821  
 /gene="SCO2257"  
 /note="Pfam match to entry PF00005 ABC\_tran, ABC  
 transporter, score 189.70, E-value 4.5e-53"  
 1297..1320  
 /gene="SCO2257"  
 /note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
 1594..1638  
 /gene="SCO2257"  
 /note="PS00211 ABC transporters family signature"  
 2174..2180  
 /gene="SCO2257"  
 2187..3011  
 /gene="SCO2258"  
 /note="synonym: SCC75A.04"  
 2187..3011  
 /gene="SCO2258"  
 /note="SCC75A.04, probable ABC-transporter, transmembrane  
 component, len: 274 aa. Similar to many transport proteins  
 including: Streptomyces peucetius  
 SW:DRRB\_STRPE(EMBL:M73758) daunorubicin resistance  
 transmembrane protein (283 aa), fasta scores opt: 523  
 z-score: 593.2 E(): 1.2e-25 35.1% identity in 245 aa  
 overlap and Streptomyces antibioticus  
 TR:Q53717(EMBL:L06249) membrane protein OLEC-ORF5 involved  
 in oleandomycin-resistance (273 aa), fasta scores opt:  
 1502 z-score: 1685.2 E(): 0.83.2% identity in 274 aa  
 overlap. Contains multiple possible membrane spanning  
 hydrophobic domains."  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable ABC-transporter, transmembrane  
 component."  
 /protein\_id="CAB61704.1"  
 /db\_xref="GI:6469243"  
 /db\_xref="GOA:Q9RKS0"  
 /db\_xref="SPTREMBL:Q9RKS0"  
 /translation="MSANYAEDTLTTPDARITPRAHURHTSALIRNLLWIRQDPES  
 MFDALFFIVFTLLFVIFGSGIGSGGQQAYVQYVIFGLMAMGMNAQGVGTGF  
 NQDFNSGVNDRFRSLPIGRSGVLFKIAVELLMFLAVLMVAVLVGFDITDWAGL  
 LASVGLSAVFGSALMWVFLTLGVNMKQSQVAMGFLVLMPLQFGSSIFADTSPGWM  
 LOSFTDYNPLSLADTARGMLGGPVADHDLVTLWSAALTAVMAPIAIHKFRTKS"  
 complement (3033..7037)  
 /gene="SCO2259"  
 /note="synonym: SCC75A.05c"  
 complement (3033..7037)  
 /gene="SCO2259"  
 /note="SCC75A.05c, possible multi-domain regulatory  
 protein, len: 1334 aa. Similar in parts to many  
 regulators. Similar in the N-terminus to Streptomyces  
 coelicolor SW:AFSR\_STRCO (EMBL:D90155) regulatory protein  
 Afar (993 aa), fasta scores opt: 524 z-score: 356.0 E():  
 1.9e-12 37.0% identity in 316 aa overlap. Similar in the  
 C-terminus to many other proposed regulators e.g.  
 Mycobacterium tuberculosis TR:O53720 (EMBL:AL021931)  
 transcriptional regulatory protein (1085 aa), fasta scores  
 opt: 809 z-score: 541.8 E(): 8.3e-23 34.1% identity in 883  
 aa overlap. Contains a Prosite hit to PS00017  
 ATP/GTP-binding site motif A (P-loop). Also contains a  
 highly degenerate repeat region with 16x PA(P/S)G and 5x  
 APG. This repeat region also has a notably higher GC  
 content (83.4%). Contains probable coiled-coils between  
 residues 867 to 904 (38 residues) and 1093 to 1121 (29  
 residues)."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative multi-domain regulatory protein."





```
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes S.E., O'Hearn E., Rosenblatt A., Callahan C., Hwang H.S.,
Ingersoll-Ashworth R.G., Fleisher A., Stevanin G., Brice A.,
Potter N.R., Ross C.A. and Margolis R.L.
TITLE A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
JOURNAL Nat. Genet. 29 (4), 377-378 (2001)
MEDLINE 21583737
PUBMED 11694876
REFERENCE 2 (bases 1 to 125020)
AUTHORS Holmes S.E., Ingersoll-Ashworth R.G., Ross C.A. and Margolis R.L.
TITLE Direct Submission
JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES
source
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/notes="isolated from a patient with Huntington's
Disease-Like 2 (HDL2)"
complement(35581..35746)
/rpt_type=tandem
/rpt_unit="ctg"
complement(336507..>36887)
/genes="JPH3"
/notes="synonym: JP3"
complement(336507..>36887)
/genes="JPH3"
/product="junctophilin 3"
complement(336507..>36887)
/genes="JPH3"
/notes="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translators="MSBGRFDDGSGVCGHEDKAGHGVCTGPKQGYTQSW
HGFEVLGVYTPSGNTYQGTWAGQKRHIGLESKRWYKGEWTHGFKRGYGVRECA
NGARYETGWSNGIQDGYGTETYSDG"
repeat_region
gene
mRNA
CDS
ORIGIN
Query Match 10.4%; Score 54; DB 9; Length 125020;
Best Local Similarity 10.9%; Pred. No. 23;
Matches 45; Conservative 220; Mismatches 143; Indels 4; Gaps 2;
Qy 44 GAATGCACTCCGGCGGGTGACCAAGTCGACACAGGTGGAGTTGAGCAGCCAGTCA 103
Db 51141 SRWYSSMAWYKGGRRKTKSMATGCMRMGAMRSKGMRYKWSCRYKGMRWGWSGY 51082
Qy 104 CGCCGCTGCGCGGATGCGGAACACCGGGCCAGGATCAGTGCAGCAGCGGATGCG 163
Db 51081 KGCSGYMSMGYSGRTSKYCSYVYGGYCSKSAAMCKSKSTSKCTKSKYSCS--MGRY 51024
Qy 164 CGCTCGATCTCGACGGTGGCGCGCAGCTGATCTCGTCCCGGGTCCACACACAGGGG 223
Db 51023 SSYCTCGWRCKWKMSSCCASRSRGSKYKMGASWRCRRGRGMSYWSMSKYSNAYK 50964
Qy 224 AAATGGCTCGGTGGCAGCGCACCCAGCCGAGCAGCTGCTCAGGCTGGCGTCACG 293
Db 50963 SSARGRGMKCYKTCMKYIYSMBKSHDSBGSYKSYRKYBYWTCMMKDSNCWSVDSHM 50904
```





Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H.,  
Isenberger, T.A., Peck, R.P., Pohlschrod, M., Spudich, J.L.,  
Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P.,  
Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and  
DasSarma, S.

Direct Submission  
Submitted (14-JUN-2000) Institute for Systems Biology, 4225  
Roosevelt Way NE, Seattle, WA 98105, USA

## FEATURES

## source

1. .14720

/organism="Halobacterium sp. NRC-1"

/mol\_type="genomic DNA"

/strain="NRC-1"

/db\_xref="taxon:64091"

complement (152..1093)

/gene="VNG1350C"

complement (152..1093)

/gene="VNG1350C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11

/product="Vng1350c"

/protein\_id="AAG19683.1"

/db\_xref="GI:10580863"

/translation="MSFRPTLGETPTTVLFLLLAAIWSSFVAIKIGIADVPALISFA

AIKYSFAGILVLAAYHTDRWRPRGRREWLAAVGLVIGIVHGAIVLGHVDVSAF

IAIVTSLAPLITAGPASKLHENVITDGTQAAGLAGAGVIVANPGASGVDLGGI

LIVFLAAAFALGGVLTGKLPGGGLPVTEWAWMLIGSLALWGAVRGENTAVH

WPTLAASLAYITLVCAGGGLFVILLERAGATENILVSYLEPIGASVVAWALPGDV

VSPRALAGFAVIFAGPALVKHRYIGAVENTTDDAEL"

1236..1721

/gene="trh6"

/note="synonym: VNG1351G"

1236..1721

/gene="trh6"

/note="trh6"

/codon\_start=1

/transl\_table=11

/product="transcription regulator"

/protein\_id="AAG19684.1"

/db\_xref="GI:10580864"

/translation="MDRDTITLKAIDLTGSPALOBETGIPVSTIHYRLNLRDD

GVINDLYINFDISIGDVTIVLVLTSDGGYEDVADEISAVEGVQTFTFMGTDF

VVARLPGGDDVERLISDEALDPVDRTNSTFVIDRHKDAQRALESYSLSLVTALAD

E"

complement (1729..3597)

/gene="gatB1"

/note="synonym: VNG1352G"

complement (1729..3597)

/gene="gatB1"

/note="gatB1"

/codon\_start=1

/transl\_table=11

/product="Glu-tRNA amidotransferase"

/protein\_id="AAG19685.1"

/db\_xref="GI:10580865"

/translation="WTFEYDELGLVAGLEIHQOLDTATKLCACPTTREPABADRT

FTYLPITSGEISDEIALESERVEFEYLADTTCUVEDEDPFPHRLDDALAAA

LEITHLLGMDVRAHVRKVIDGNTGQFQRISPAAGGAIDTSEGVGIEDLMLE

ESQAIRIEDRXYSLDLGLPLVIGTKPDISPAQAREARERIGMLRSTGVK

RGLGTIRQDVNSIADGARVEMKGVQSLDIDLVREVRQVELLDIVDELDDADAA

VGPRTVDTESGIVTGLDDGGEVHAPLHGFDGLVGRQLQADRLRGLTFPSDH

ATRHGAGGIPTHTDELPAVGTVAAEVALRDVAGGDDAVAIVADDPETAQSIQAVA

ERATAMAGVPESTRANDGTGSKYLRPLPGARMPYPTDVPDPDSAVETPELLT

EKVRYQADFLDLGAEQVAYRQWLFQEQVAGVDATLAAQLLESTVTSLRDDV

FVGLRDLHFRVGLGVADGLAQEGVPELLAAEPQGSDFAVLAELGLGSAADE

VREAVGVVERNSDQVAAGMGFSAIMCEMGALRGKADGLVSEVLRBEIQQRs"

complement (3634..5178)

/gene="VNG1353C"

complement (3634..5178)

/gene="VNG1353C"

/note="conserved hypothetical protein"

/codon\_start=1

/transl\_table=11  
/product="Vng1353c"  
/protein\_id="AAG19686.1"  
/db\_xref="GI:10580866"  
/translation="MKLHPLSVPRALSRAIFGMAFLIGGAALSGNAALTCGAVLGV  
VMVAVVAVYETAYVREDYELTEDSLDIASGVFSREREIPLRVQNVDTVRSIVA  
RGLGLAVDVETAGCGSTASLRFVSREAAADLQADVTRTRTPGDAADAATAATDA  
GGADSSRSRAADREPTPVGERGELLFALSNRDLGLYLLSFPNPLFGSIIAIVATA  
APSLGGMVDFDGMVVLVAGAVALGVMVLSAAVVRVQFGFRLLRRVGGDDURVBRG  
LFRDRGTTPLSKLTQVSYVEDSVIMERYGFAVLAETAGYAPQSGSGSEAAVPIAP  
RAAVFLAREIEPVADIDLSRPPERAQDRYRYALAGLVVAAAGVVAAGVSLRLVVSFPWY  
ALAVLPLVAPAAARAHATRGVYSDSDHVVTQAGWWRKTIVVPAHRVOTLVKQTVF  
QREWALTSVIDTAGSRSLSGGGAVALDRDGGADALADDVDDTLLSGVGRGDDTAA  
AGSAAD"  
complement (5175..5654)  
/gene="VNG1355H"  
complement (5175..5654)  
/gene="VNG1355H"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="vng1355h"  
/protein\_id="AAG19687.1"  
/db\_xref="GI:10580867"  
/translation="MERLDARVVVWMLRVFVAVVLAGAAGTGGALDVGVSAAAA  
VAVGVVALLGATHALLRYRIMRFDVREDTLYIEHGLVVRVTRTPVVRVQHVDSRRG  
PLERVLGLGRVVVYTAGSGADVTIPQLSRERAADAQETLALAISEPESGGDVAV"  
complement (5699..7111)  
/gene="fumC"  
/note="synonym: VNG1356G"  
complement (5699..7111)  
/gene="fumC"  
/note="fumC"  
/codon\_start=1  
/transl\_table=11  
/product="fumarate hydratase"  
/protein\_id="AAG19688.1"  
/db\_xref="GI:10580868"  
/translation="MSEDYRTEQDSLGEWQVPADAVGQAOTQRAIENFPISGIAFGRR  
FVRALGVKAAAQNRDLGVDDEADAIVAAADEVIAAGEHDDQFPDVFQSGSTS  
SNMNAEVIANRAEELGEBIGRVRVHNDVNVNGOSSNDVIPTAMHVASLDALVNDV  
KPGLETLAELDKADAFDGVKVTGRTHLQDAPVRLGQFGVGRVQVEKIGDIRIEAV  
APRLSELALGTVAGTGLNTHPEFTAGYISEETGVTFREADNHEFAQAADANNE  
AHGRLTGVGNLKIANDRLKSLASFRANGLGEIQEQPNQSSIMFGKINPVVAEAVN  
QVHKVQVGNDAALAAAGAGQIDINLYKPVLAHNFLOSADMLANASAAAFGEKFAKLE  
ANEAACEAQRVSMALATALNPTIGYDKASEVAKAALKEGKTIVTEVVVEKYLSEABA  
ADVLDPGEMTHRGILSGDDT"  
complement (7315..7563)  
/gene="VNG1357C"  
complement (7315..7563)  
/gene="VNG1357C"  
/note="conserved hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="Vng1357c"  
/protein\_id="AAG19689.1"  
/db\_xref="GI:10580869"  
/translation="MLADIEARIEAGFEADCSATVSRARGPHDDHLEAVVVPAPF  
GVSLVDQHDVYDALGDAMTEITHALELITRPSDRDA"  
complement (7612..8595)  
/gene="rad2"  
/note="synonym: VNG1359G"  
complement (7612..8595)  
/gene="rad2"  
/note="rad2"  
/codon\_start=1  
/transl\_table=11  
/product="DNA repair protein"  
/protein\_id="AAG19690.1"  
/db\_xref="GI:10580870"  
/translation="MGNADLRQAAIETETPFADLEGSVAVDAHNMILYKYLTTTVOVT  
GADVYTTSDGTEVANLVGAVQGLPKPKEFHEGLITPVFVWDGVTGELKDLADREQR  
YEQLDDAREAGAAEAARADARTQRLTPTTIHETTRFLLDLPQVEAPAEQAQA



YMTRTDADVAGSDYDVCILLGSPVTLRLQTSSTGHPPELMDFDATLAEHDLTWEQLVND  
VGLTCTDENPGIDTGFPTTALDAIGEHGDLMDVLAAGEHVAHGDRIRLEFLNPVDT  
DDVVDVSPDADARAFAVTDWEVDADAVERFERIDAAAQTLGDRWT"  
/gene="VNG1360H"  
/complement (8644.. .9138)  
/gene="VNG1360H"  
/complement (8644.. .9138)  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="vng1360h"  
/protein\_id="AAG19691.1"  
/db\_xref="GI:10380871"  
/translation="MNTGKAVAYEGAPLAKRGDDWPPDARAADDVDDVGVAVSFNE  
DRDAAALRYVTGDRGAGIGARLCAFAAAHLLGDECVRIAVNNPPAYEALHK  
AGFGTQETGIAELVLERCEHRAAGYQDLDRYERDLSPEEAFLDRGASPPA  
RVSW"  
9315.. .9716  
/gene="VNG1362H"  
/complement (8644.. .9138)  
/gene="VNG1362H"  
/notes="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="vng1362h"  
/protein\_id="AAG19692.1"

gene  
CDS

Query Match 10.2%; Score 53.2; DB 1; Length 14720;  
Best Local Similarity 48.7%; Pred. No. 52;  
Matches 145; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 2 GTGCAGCAGCAACACCGGCTGTCGCCCGCGCCCATCACGAACCTGCCACTCCGCGCG 61  
Db 3934 GTACACGCCCCGGTGCCTGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3993

Qy 62 GGTGACCAAGTTCGACCGAGTGGAGTTGAGCAGCAGCAGTATCGCCGCTCGCGCGCGCAT 121  
Db 3994 GGCCAGCGCTGACACGAGGACGACGACGAGCGCGCTCACAGCGCGCGCGCGCGCG 4053

Qy 122 GCCGAACACCGCGCGAGGATCAGTGCAGCAGCGCGCGCGCGCGCGCGCGCGCGT 181  
Db 4054 CACGCCACG 4113

Qy 182 CGGCGCGAGCTCGATCTCCTCCCGCGCTCCACACAGCGGGAACCTGGCTCGGTGGCAG 241  
Db 4114 CCGGAGAGTTCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4173

Qy 242 CGGACG 299  
Db 4174 CCGGCGCGGATCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4231

RESULT 33  
AP004323  
LOCUS  
DEFINITION  
ORIGIN

AP004323 95209 bp DNA linear HTG 21-MAR-2002  
OJ1118\_C02, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
AP004323  
HTG; HTGS PHASE2  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
1  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC  
clone: OJ1118\_C02  
Published Only in Database (2001)  
2 (bases 1 to 95209)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission

JOURNAL Submitted (14-NOV-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kannondai  
2-1-2, Teikuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
COMMENT The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and RGP-Japan sequencing data.  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source  
1. 95209  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="6"  
/clone="OJ1118\_C02"

ORIGIN  
Query Match 10.2%; Score 53.2; DB 2; Length 95209;  
Best Local Similarity 45.4%; Pred. No. 32;  
Matches 178; Conservative 0; Mismatches 212; Indels 2; Gaps 1;

Qy 107 CGCTTCG 166  
Db 69303 CGACCG 69362

Qy 167 CTCGATCTCGAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 226  
Db 69363 GGGGGCG 69422

Qy 227 CTGCGTCGTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 286  
Db 69423 GCG 69482

Qy 287 CTCGCGCGCGTCG 346  
Db 69483 CCG 69542

Qy 347 AGTCCCGCTGCTCCCGCTTGGCTGGAGGATAGCGGTTACGACGAGCGCGCACCG 406  
Db 69543 CCG 69602

Qy 407 GGGCG 466  
Db 69603 CCG 69660

Qy 467 CGGTGCGCGAGATCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498  
Db 69661 CCG 69692

RESULT 34  
AE016920  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE

AE016920 305584 bp DNA linear BCT 14-OCT-2003  
Chromobacterium violaceum ATCC 12472 section 11 of 16 of the  
complete genome.  
AE016920 AE016825  
AE016920.1 GI:34104104  
Chromobacterium violaceum ATCC 12472  
Chromobacterium violaceum ATCC 12472  
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
Neisseriaceae; Chromobacterium.  
1 (bases 1 to 305584)

# AUTHORS CONSRMT TITLE JOURNAL MEDLINE PUBMED REFERENCE

Brazilian National Genome Project Consortium  
The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability  
Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)  
22982980  
14300782  
2 (bases 1 to 305584)  
Vasconcelos, A.T.R., de Almeida, D.F., Almeida, F.C., de Almeida, L.G.P., de Almeida, R., Gonçalves, J.A.A., Andrade, E.M., Antonio, R.V., Araribe, J., de Araújo, M.F.F., Filho, S.A., Azevedo, V., Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., van der Berg, C., Blaney, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A., Brocchi, M., Buriy, H.A., Camargo, A.A., Cardoso, D.D.P., Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M., Cavada, B.S., Chueire, L.M.O., Fasa, I.B.C., Duran, N., Fagundes, N., Falcao, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S., Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R., Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A., Gonçalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grissard, E.C., Guimarães, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J., Leão, L.C.T., Passarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P., Macedo, W.S., do Mauro, S.M.Z., de Medeiros, S.R.B., Meisener, R.V., Martins, W.S., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F., Menck, C.F.N., Oliveira, S.C., Paixão, R.F.C., Parente, J.A., Oliveira, J.G., Oliveira, S.C., Paixão, R.F.C., Parente, J.A., Pedrosa, F.O., Pena, S.J.D., Perreira, J.O., Perreira, M., Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Petrich, D.P., Neto, C.E.R., Reis, A.M.M., Rigo, L.U., Rondinelli, E., dos Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Seunear, H.N., Silva, A.M.R., da Silva, A.L.C., Silva, D.M., Silva, R., Simões, I.C., Simon, D., Soares, C.M.A., Soares, R.B.A., Souza, E.M., Souza, K.R.L., Souza, R.C., Steffens, M.B.R., Steindl, M., Teixeira, S.R., Urmenyi, T., Vettore, A., Wasseem, R., Zaha, A. and Simpson, A.J.G.  
Direct Submission  
Submitted (22-JAN-2003) Labinfo, LNCC - Laboratório Nacional de Computação Científica, Rua Getúlio Vargas 333, Petropolis, RJ 25651070, Brazil  
FEATURES  
source  
1. 305584  
/organism="Chromobacterium violaceum ATCC 12472"  
/mol\_type="genomic DNA"  
/strain="ATCC 12472"  
/db\_xref="taxon:243365"  
370. 2154  
/locus\_tag="CV2796"  
370. 2154  
/locus\_tag="CV2796"  
/notes="identified by sequence similarity; putative; ORF located using Blastx/COG0006"  
/codon\_start=1  
/transl\_table=1  
/product="probable peptidase, M24 family protein"  
/protein\_id="AAQ60464.1"  
/db\_xref="GI:34104105"  
/translation="MTIPQRTASFAAMKAGADACLVPSDDPHISYLPQWQARQ WLSGFTGSMGLTVGADDFAGLWADSRVYQAEQELAGTGIQLMKIDTAASSLHLOWLA DNRCTQTLVDGDLGLAAKALQALPGRGVIVRADLDDLDVAVAGRPALPAAYV QHDDTAPSRRAKLARVREMGADRHFIITLDDIWLNLGAGVSYNVPFIH ALIETDNATLFWAPGKIDIALAQLAAGDVRADYAAKPAALSPALSGALLIDPRVT LGURVADRVRIEINPTSLMKRSAAEADHVRAMEQDGAALAEFFAFVFNVN RARITLITDDEITAAARPPGVSPFSGTIFAGNANGALPHYATPEHSEIQDGL LITDGGVGLGTTDITRVVAVGTPSAOKRDTFLVLTGKTMALSMHAFPRGLTSMELD ALARALPWHDIDFGHTGTVGVFLNVHGPQISRAVPEPHMAQEGMTSIEPGV YRAGONGVRIENLVANVSPRNQPGFLKPELTLCPIIDRCIDLSLLTQPEIAMNG YHAEVRARLPLVDGARDWLLANTETL"  
2225. 3118  
/locus\_tag="CV2797"  
2225. 3118  
/locus\_tag="CV2797"  
/notes="identified by sequence similarity; putative; ORF located using Glimmer/GeneMark/Blastx/COG0583"  
/codon\_start=1

/transl\_table=1  
/product="probable transcriptional regulator, LysR family"  
/protein\_id="AAQ60465.1"  
/db\_xref="GI:34104106"  
/translation="MPFRLADRLRFQDASQGSFSAARXKHLSPAAASACIQRMBA ALNARLFTTRRLRAGETFLAYCQALEEGRHLQQAQCSGLVRLSAPS DLRNRLLDCLDRFGAHPAVHSLSDTPADLLICDDIDLAIRYQCPADSLVARQL RASRVCAAPSLLEKLGQSHPODLAGLPTLTLTGHPNMWRYRUKGELRSRLLE RYQNSDGEVLERWAVQGHGFAVKSELDIADLAAGRLTTVLDLDDYFIEBAPLHLLYPG HRLQPARIRLLIDFLIDQPA"  
complement (3474. 4790)  
/locus\_tag="CV2798"  
complement (3474. 4790)  
/locus\_tag="CV2798"  
/notes="identified by sequence similarity; putative; ORF located using Blastx/COG0477/TC:2.A.1.21.1"  
/codon\_start=1  
/transl\_table=1  
/product="probable macrolide efflux protein"  
/protein\_id="AAQ60466.1"  
/db\_xref="GI:34104107"  
/translation="MGAAFAFCMLSEMITGLSGLLQLFALGVNIYQKTSVEDFSFAL LGTLPAFILLPIAGACADLLDRRKVLIIACDFLLILLIGVLSFMILWDLALVPAAMAG NGVALIHLPFRSPVQVALSRLLPUNQSOANGRLTSLGLVQLAGPLMAGALMFOFG LITGLPTELLIARAVLLVLRMGRNTEPAAAGNTVGLRWRHSRSLSKTMYALNGQR AMFGLLIYMLLQALITLVSSMITPLVLRHGSDDTIGVMVGGCAFCAIGATVITISN PERLMRWIISDLALAVAMAGFVEKFLWALAAAMACGSVNSCTLSLWMEKT PLDRGSLFSLIASINLAASAMLAGVILVERVLAPAMVPGGLADLLGLWIGTQGG RGVALLVLCGIGLSAISVFALLGTNLALQDDMVEDAPICLSESA"  
complement (4817. 6136)  
/locus\_tag="CV2799"  
complement (4817. 6136)  
/locus\_tag="CV2799"  
/notes="identified by sequence similarity; putative; ORF located using Glimmer/GeneMark/Blastx/COG3208"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAQ60467.1"  
/db\_xref="GI:34104108"  
/translation="MTVKRMVQDVSVGSAALASLEGEBTGRWLAAQPEHGQTQOG LAYALAEVRLQDPSFYRLAGVGLAYEVLQRLGMQVTCVACLEGAPAWVRQ SEALAVERSQALSPAWLRCLSSVYPEPGDFLPQNTMPASAEAFWRALGAQASARVP AHQPLVMRGROSLPPLICPGAGASVVDFFALSSALYQGSVYGLQPRGLDROPP HVTVEDAARSYLDALANAPGVHLLGHSGFWIAHSLAAQLQASQATRVLSLTVLDSR VPNDAGEVQOREVLSELIALYQGGEGEPLRLDAELASREOSKALHGLLELLYGI LPSRSSPDILSGVIRTFACALRAGYRPERLYQGSRLVLVKPSEASSEATSEATSEQWI AGRRRLPDLVFWEGFNGNHITLRSPOAGHLAAWLDMLFDMKAVHS"  
complement (6199. 6627)  
/locus\_tag="CV2800"  
complement (6199. 6627)  
/locus\_tag="CV2800"  
/notes="identified by sequence similarity; putative; ORF located using GeneMark/Blastx"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AAQ60468.1"  
/db\_xref="GI:34104109"  
/translation="MALRMGRPAGRLPGVWALGILLASACVGNKLEEDVDPEYVKQ ALSRGETKQMDLFCQPORKISFDGDNEIWTYSYRQYKAKLQNLTNVSVLRGQNN YSKELVLFDSKVRKRWKLKASEESVGTGFATQRPAN"  
complement (6644. 7489)  
/locus\_tag="CV2801"  
complement (6644. 7489)  
/locus\_tag="CV2801"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAQ60469.1"  
/db\_xref="GI:34104110"  
/translation="MRSAGFGLAVAMLTSAALAGTAADLASSRTYAVLQKGAVFQFE NAKLAPNGVYQYRLYLINGVSNASAPTLKLSRGGSKAVPLGARQVFLSPGTAEKSDAS

[illegible]

```
misc_feature 82587..84446
/note="ORF 29; positive strandedness"
misc_feature 84481..85548
/note="ORF 30; positive strandedness"
misc_feature 85556..86845
/note="ORF 31; positive strandedness"
misc_feature complement(86803...87372)
/note="ORF 32; positive strandedness"
87494..88420
/note="ORF 33; positive strandedness; N-terminus only"

ORIGIN
Query Match 10.2%; Score 52.8; DB 6; Length 88421;
Best Local Similarity 46.5%; Pred. No. 38;
Matches 240; Conservative 0; Mismatches 272; Indels 4; Gaps 2;

Qy 7 GCACGAACGACCGGTGTGCGCCCGCGCCATCAGAACTGCCACTCCGGCGGGGTGA 66
Db 67518 GCACGAGCAGTGGCGCCCGCGCGGATGAGCGCGCTGCGCGCGAGCGCA 67577
Qy 67 CCAGGTGACAGGTGGAGTTGAGCAGCAGCTATCGCGCGCTGCGCGCGGATCCGA 126
Db 67578 ACTCGCCGCGCGTGGCGCGGAGCGCGGTGCGCGCGTGTCTGCTCCCTACCGGA 67637
Qy 127 AACACCGCGCGCAGATCACGTGACGACGCGATGCGCGCTCGATCTCGACGTCGGCC 186
Db 67638 CGGCTCGCGCGACCTGTGTGTGGCCCGCGACCGCTGACCGCGACCGTGTAT 67697
Qy 187 GCAGCTCGATCTCTGCCCGGTCTCCACACGAGGGAATCTGGCTGCTGCGCGGCA 246
Db 67698 CGCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67757
Qy 247 GCGCCGCGCGGACAGCTGCTGACAGGCTGCGCTCAAGGCTCTCGCGGTCGCGGCG 306
Db 67758 GCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67817
Qy 307 AACCGCGCAGCGCTCAGATCCCTGTAGTCGCATCGCTCAGTCCCGGTGCTCCCTTG 366
Db 67818 CGAGCTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67875
Qy 367 GCTTGGGAGGATACGCTTTCAGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 424
Db 67876 GCGAGGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67935
Qy 425 TTGAGCGCATCGCTCGATGACACGCGCGCTCGCGGTCGCGGCGGTGCGCGGATCCGTA 484
Db 67936 TTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67995
Qy 485 CCGCGCGGACCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 520
Db 67996 GCGCGCGGTCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 68031

RESULT 36
AC069466/c 192634 bp DNA linear HTG 06-SEP-2000
LOCUS Mus musculus clone RP23-394J16, WORKING DRAFT SEQUENCE, 23
DEFINITION unordered pieces.
AC069466
VERSION AC069466.2 GI:9972319
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 192634)
AUTHORS McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Dedhia,N.N., de la
Bastide,M., Huang,E.N., King,L., Kirchoff,K.A., Miller,B.,
Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Rodriguez,M.A.,
Shah,R.S., Shekher,M., Spiegel,L.A., Toth,X. and Vil,M.D.
MUSE Genomic Sequence
Unpublished
TITLE Mouse Genomic Sequence
JOURNAL
REFERENCE 2 (bases 1 to 192634)
```

AUTHORS  
TITLE  
JOURNAL

McCombie,W.R.  
Direct Submission  
Submitted (01-JUN-2000) Lita Annenberg Hazen Genome Sequencing  
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring  
Harbor, NY 11724, USA  
On Sep 6, 2000 this sequence version replaced gi:8134854.  
----- Genome Center  
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor  
Laboratory  
Center code: CSHL  
Web site: http://www.cshl.org/genseq  
Contact: mcombie@cshl.org  
----- Project Information  
Center project name: RP23-394J16  
Center clone name: RP23-394J16

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 30565: contig of 30565 bp in length  
\* 30566 30917: gap of unknown length  
\* 30918 52496: contig of 21579 bp in length  
\* 52497 52848: gap of unknown length  
\* 52849 71268: contig of 18420 bp in length  
\* 71269 71620: gap of unknown length  
\* 71621 83193: contig of 11573 bp in length  
\* 83194 83545: gap of unknown length  
\* 83546 93802: contig of 10257 bp in length  
\* 93803 94154: gap of unknown length  
\* 94155 103828: contig of 9874 bp in length  
\* 103829 104180: gap of unknown length  
\* 104181 113354: contig of 9174 bp in length  
\* 113355 113706: gap of unknown length  
\* 113707 121968: contig of 8262 bp in length  
\* 121969 123200: gap of unknown length  
\* 123201 129192: contig of 7072 bp in length  
\* 129193 129744: gap of unknown length  
\* 129745 136687: contig of 6943 bp in length  
\* 136688 137039: gap of unknown length  
\* 137040 142803: contig of 5764 bp in length  
\* 142804 143155: gap of unknown length  
\* 143156 147963: contig of 4808 bp in length  
\* 147964 148315: gap of unknown length  
\* 148316 153054: contig of 4739 bp in length  
\* 153055 153406: gap of unknown length  
\* 153407 157827: contig of 4421 bp in length  
\* 157828 158179: gap of unknown length  
\* 158180 162507: contig of 4328 bp in length  
\* 162508 162859: gap of unknown length  
\* 162860 167154: contig of 4295 bp in length  
\* 167155 167506: gap of unknown length  
\* 167507 171735: contig of 4229 bp in length  
\* 171736 172087: gap of unknown length  
\* 172088 175937: contig of 3850 bp in length  
\* 175938 176289: gap of unknown length  
\* 176290 180015: contig of 3726 bp in length  
\* 180016 180366: gap of unknown length  
\* 180367 184069: contig of 3703 bp in length  
\* 184070 184420: gap of unknown length  
\* 184421 187865: contig of 3445 bp in length  
\* 187866 188216: gap of unknown length  
\* 188217 191451: contig of 3235 bp in length  
\* 191452 191802: gap of unknown length  
\* 191803 192634: contig of 832 bp in length.

Location/Qualifiers

1..192634  
/organism="Mus musculus"  
/mol\_type="genomic DNA"

## FEATURES

source



gene	WPRG"		IDALLAGYRVFIEVSPHPVLIPALRECFEAEVAAATVPTLRDQGGPPQDQVVARALGD	
	complement (6768. .7493)		GVFAGLAVDWSRWFVGDREAGDEGHRPRTVELPTYPQRRRYVLAADPHGREGKRTAG	
	/gene="nanr1"		VTRPAGHALLUSSAVELADGGLVLSGRLLPGDAAWVGAHTVAGVQLVPGAVLVDWALLA	
	complement (6768. .7493)		ADRAGASLEILLRLAPLELSGPSAGVLAQVAVCAPDESGRERLURISRPAD	
	/gene="nanr1"		AGAGEGWTCHAVGSLAPGPPADTGTATVPHPPAGAEALDPAGLYERERGVYGV	
CDS	/notes="putative regulatory protein"		FAURGVVALWRDGLADLVADVALPEERAGGEGGADGDTAGFGLHPVLLDAALQFALL	
	/codon_start=1		AEPDGTGEGAGPEARLWLFPAWSGVRLWATGARAARVRLSPLDGGGSDVADERELRI	
	/transl_table=11		EVSDDTGAFLVLSVVLRRPTVRQVREASGAAGLFDALDFTVPAPQEPESGADDDAG	
	/product="Nanr1"		CVALGEPATPEPGVDCRDTYTDLPALLAALDAGAPLPSVMMRPAAADPCGAPEDAA	
	/protein_id="AAP42853.1"		LSAVRGVAAALRAWVAEPRLLTVSRLLAVVTRGVAAGAGGEPVDLAAAGAAWGCARGVQ	
gene	WGLPFRRTASIVICISLRKFLSPGQNNPIITRPGGYMLRRASBSCDFVFOQLV		AQUHMTVDRDRTHALERVVAQVDPAPHLTGUVHVAGLSADIETSGRAGWAGVAAGVU	
	ADKRCVYGEPAQALCFEQALALWRFAGFNVGTGPIVAGFLKRLDEIRTECAEMS		RFGWAAADDDGAGSGRGTGVLGADRGALLRAACRADRPRLVAADIRTRALTAHPA	
	AEAHLELGRHLELVGRLSLVAEFLRETFFYRLQMLALYRSQKSEALGIYESARRVL		HELPAALRTLLGASASASAGGRAPVRRVAAAAAPGRTTDMASRLVGLGPAERRRAVLELV	
	DRELGPGRKRLRELARTVLAST"			
	complement (7573. .8334)			
CDS	/gene="nanr2"			
	complement (7573. .8334)			
	/gene="nanr2"			
	/notes="putative regulatory protein"			
	/codon_start=1			
gene	LVKEIWTGTRAPQAVAAHLYVSQLRFLRPNRTSPITAKSGVYLRIGDDELDFH		LVKEIWTGTRAPQAVAAHLYVSQLRFLRPNRTSPITAKSGVYLRIGDDELDFH	
	VFQGLVQGAAGAGCTHEACEAYEEGLEFWRGPLYLEDLDGAIYNGFVWAEQARL		VFQGLVQGAAGAGAGCTHEACEAYEEGLEFWRGPLYLEDLDGAIYNGFVWAEQARL	
	ECVESFVEAGLELGRHREFVSYLARIELPLNEAFYRLQMLALYRSQKSEALGIYESARRVL		ECVESFVEAGLELGRHREFVSYLARIELPLNEAFYRLQMLALYRSQKSEALGIYESARRVL	
	SARAVLSEELGLEPSAALKRLHNAVLVDAPELEAA"		SARAVLSEELGLEPSAALKRLHNAVLVDAPELEAA"	
	8919. .17627		8919. .17627	
gene	/gene="nanA1"		/gene="nanA1"	
	8919. .17627		8919. .17627	
	/gene="nanA1"		/gene="nanA1"	
	/note="PKS"		/note="PKS"	
	/codon_start=1		/codon_start=1	
CDS	/transl_table=11		/transl_table=11	
	/product="NanA1"		/product="NanA1"	
	/protein_id="NanA1"		/protein_id="NanA1"	
	/db_xref="GI:31044143"		/db_xref="GI:31044143"	
	/translation="MRRFYGVWAGSSPSHAQOATSPVAIVGLACRLPGADPEAFWR		/translation="MRRFYGVWAGSSPSHAQOATSPVAIVGLACRLPGADPEAFWR	
gene	LLRAGNAVVPDPSRTPFEGSPYFAGLLEHVDTPDAGFFGISPREATAMPQORL		LLRAGNAVVPDPSRTPFEGSPYFAGLLEHVDTPDAGFFGISPREATAMPQORL	
	MLELAWEMEDAGLGNLABERTAVFTGAIWDYATLLHRRPNIDIAITRHTMAGLH		MLELAWEMEDAGLGNLABERTAVFTGAIWDYATLLHRRPNIDIAITRHTMAGLH	
	RGLIANVRHLLGDLGSLTVDAQAQSSGLVAVHLACBSLRGRBADLALAGVNLILAE		RGLIANVRHLLGDLGSLTVDAQAQSSGLVAVHLACBSLRGRBADLALAGVNLILAE	
	ESMWAQAFQGLSPDGRCTFDARANGFVRGEGGMVLLKPLAAAVADGDPVYCVIE		ESMWAQAFQGLSPDGRCTFDARANGFVRGEGGMVLLKPLAAAVADGDPVYCVIE	
	GSAVNTGADTGLTRPSADACTVRCAMORAAVSFAELQYVELHGTGTPVGPPIEAA		GSAVNTGADTGLTRPSADACTVRCAMORAAVSFAELQYVELHGTGTPVGPPIEAA	
gene	ALGDALGDGDCRRDRAVDAPLAVGSKTVNGHLEAAAGIAGLLKTLAISIHRRLLP		ALGDALGDGDCRRDRAVDAPLAVGSKTVNGHLEAAAGIAGLLKTLAISIHRRLLP	
	SLNFAITNPGLPLAELGLRVDTAFGPMDBERRRLTAGVSSFGMGNTNCHVILAEPPAP		SLNFAITNPGLPLAELGLRVDTAFGPMDBERRRLTAGVSSFGMGNTNCHVILAEPPAP	
	AVLDPRTPAGDVSARTAPPVMPVWSAASPALTAQAAALYELHRAHPGLHFDVIGH		AVLDPRTPAGDVSARTAPPVMPVWSAASPALTAQAAALYELHRAHPGLHFDVIGH	
	ALATRTAFPHRAVVLGRDEDELVSRLDALASSETQTSVIRAGAGRVAFPLFSGQGS		ALATRTAFPHRAVVLGRDEDELVSRLDALASSETQTSVIRAGAGRVAFPLFSGQGS	
	QRPGWRELVAAYPVFADALREVCAHLDPMLDTDPLDLMFAEAPDGPBPPLNRTAY		QRPGWRELVAAYPVFADALREVCAHLDPMLDTDPLDLMFAEAPDGPBPPLNRTAY	
gene	TOPALEALVAYLRLVTSWGTVPDHLMGHSGEITAAHVAGVLSLPDACLVAARGKL		TOPALEALVAYLRLVTSWGTVPDHLMGHSGEITAAHVAGVLSLPDACLVAARGKL	
	MOSITAPGMAAMQATAEAGQALVAVGVRGLAAVNA PASVVISGDR EAVAEATAAW		MOSITAPGMAAMQATAEAGQALVAVGVRGLAAVNA PASVVISGDR EAVAEATAAW	
	RARGKATLVKVSHPSPHLDGILGLRTVVAAGLTPAAPAI PVWSNLTGGA TEATLQ		RARGKATLVKVSHPSPHLDGILGLRTVVAAGLTPAAPAI PVWSNLTGGA TEATLQ	
	RSPDYADHARQAQVRFADGVGRHLCDAGVDITFLELPGDASLTGMARSAAGAWGDAFYP		RSPDYADHARQAQVRFADGVGRHLCDAGVDITFLELPGDASLTGMARSAAGAWGDAFYP	
	VAVAVQRGRPEAGFSVSAQAQHVGRGVGDVMAAFAGHETRAPLFTYAFQDRHWP		VAVAVQRGRPEAGFSVSAQAQHVGRGVGDVMAAFAGHETRAPLFTYAFQDRHWP	
gene	DGLDERGARRESTSPVPESTDREPVVDVADSPADPAASPGELLALVTRHAALVIGHNSP		DGLDERGARRESTSPVPESTDREPVVDVADSPADPAASPGELLALVTRHAALVIGHNSP	
	DGIDPALTQKQGLFDSLAATBELSRLSAAITDELPTATITFDHPTPNAVAWLAHAEHG		DGIDPALTQKQGLFDSLAATBELSRLSAAITDELPTATITFDHPTPNAVAWLAHAEHG	
	OPTAAPTAAFGSMAEDPVAVVAVSCRYPGGVSEALWELVDGVDAVGEFFCDRGM		OPTAAPTAAFGSMAEDPVAVVAVSCRYPGGVSEALWELVDGVDAVGEFFCDRGM	
	DLAELFGAPDSGSGATRGVFLYAGDPAEFPGISPREALAMDPOQRILLLELSNE		DLAELFGAPDSGSGATRGVFLYAGDPAEFPGISPREALAMDPOQRILLLELSNE	
	LLEBAGIPASLASAGTGVYATADYGPRLHEATAELDQHLITGSPBVSAGRVAY		LLEBAGIPASLASAGTGVYATADYGPRLHEATAELDQHLITGSPBVSAGRVAY	
gene	ALGLEPCTVDTACSSSLVNHLLAAQALRQCECDLALAGGVTWATPGNFTSFQR		ALGLEPCTVDTACSSSLVNHLLAAQALRQCECDLALAGGVTWATPGNFTSFQR	
	GLADPGRCPPAARADGTGSEAGLVLLERLSDARRNGHQLAVIRGSVAVNQDGSN		GLADPGRCPPAARADGTGSEAGLVLLERLSDARRNGHQLAVIRGSVAVNQDGSN	
	GLSPNGPSQQRVIRQALANARLEPDAVVAEHGTGTGLGDPTEAQAALLATYGGQRT		GLSPNGPSQQRVIRQALANARLEPDAVVAEHGTGTGLGDPTEAQAALLATYGGQRT	
	DDRLPLWGSIKSNTGHTQAAGVAVGI MMVNALRHRLPASLHIDAPSHPIDSDGT		DDRLPLWGSIKSNTGHTQAAGVAVGI MMVNALRHRLPASLHIDAPSHPIDSDGT	
	RLLSBPVMGPDPGSPRRAAVSFGISGTINAHILTEQAPDHPPEPTTSGGVYV		RLLSBPVMGPDPGSPRRAAVSFGISGTINAHILTEQAPDHPPEPTTSGGVYV	
gene	VLSRTADALRQAQGLAELWATAGAPSPASPTSPASPADVGHSLATSTRASDRHRAV		VLSRTADALRQAQGLAELWATAGAPSPASPTSPASPADVGHSLATSTRASDRHRAV	
	VSGTRDELISGLRADGLAPAAVSAGAAPGVPMVPPGSGWRMGVLLDSSPVF		VSGTRDELISGLRADGLAPAAVSAGAAPGVPMVPPGSGWRMGVLLDSSPVF	
	NARMAEAAALGFEVDVSLTAVLRGAPGAPERSVDVLQPCLMWYSLAAVWESYCV		NARMAEAAALGFEVDVSLTAVLRGAPGAPERSVDVLQPCLMWYSLAAVWESYCV	
	TPTAVVGHSGEIAACVAGGLSRDGRARVLRSLQALRALLAGHTWASLALSGAER		TPTAVVGHSGEIAACVAGGLSRDGRARVLRSLQALRALLAGHTWASLALSGAER	
	RFLADLGAARAVTVFNGPSTVVSGETDQDVAAVAAEAAAGHRTATIDVDVASHG		RFLADLGAARAVTVFNGPSTVVSGETDQDVAAVAAEAAAGHRTATIDVDVASHG	
gene	POVDRLATITRDLADLSPGASDAVFYSAVTGARQPTTELDADYTNLRQVPRFASA		POVDRLATITRDLADLSPGASDAVFYSAVTGARQPTTELDADYTNLRQVPRFASA	

AC135396 134544 bp DNA linear HTG 13-DEC-2003  
Medicago truncatula clone mth2-33018, WORKING DRAFT SEQUENCE, 2  
unordered pieces.  
AC135396  
AC135396 GI:39841085  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS  
Medicago truncatula  
SOURCE  
ORGANISM  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE  
1 (bases 1 to 134544)

AUTHORS	Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
TITLE	Cook, D., Kim, D. and Roe, B.A.
JOURNAL	Medicago truncatula BAC Clone mth2-33018
REFERENCE	2 (bases 1 to 134544)
AUTHORS	Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE	Direct Submission
JOURNAL	The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
REFERENCE	3 (bases 1 to 134544)
AUTHORS	Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D. and Roe, B.A.
TITLE	Direct Submission
JOURNAL	Submitted (13-DEC-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA
COMMENT	On Dec 13, 2003 this sequence version replaced gi:39752708. ----- Genome Center Center: Department Of Chemistry And Biochemistry The University Of Oklahoma Center code:UOKNOR
FEATURES	* NOTE: This is a 'working draft' sequence. It currently * consists of 2 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. *       1     2202: contig of 2202 bp in length *       2203   2302: gap of unknown length *       2303   134544: contig of 132242 bp in length. Location/Qualifiers 1..134544 /organism="Medicago truncatula" /mol_type="genomic DNA" /db_xref="taxon:3880" /clone="mth2-33018" /clone_lib="Medicago truncatula BAC library H2"
ORIGIN	
Query Match	10.1%; Score 52.6; DB 2; Length 134544;
Best Local Similarity	45.1%; Pred. No. 37;
Matches 214; Conservative	0; Mismatches 259; Indels 2; Gaps 1
Qy	42 ACGAACTCCACTCGCGGGGTGCACAGTTCAGCAGGTGGAGTTGAGCAGCCAGCTC 101
Dn	307 AGNNNNNNNNNGGG 366
Qy	102 ATCGCGCGCTGCGCGCGCATGCCGGAACAACCGGGGCAGGATCACGTGACGACGCGCATG 161
Dn	367 GGGGCGGCGGGGCGCGGGGGCGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGGG 426
Qy	162 CGCGCCTCGATCTCGAAGGTTCGGCGAGCTCGATCTCGTCGCCGCGCTCCACACCAGG 221
Dn	427 CGCGGCGGCG--CCCCGCGCGCGCGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCG 484
Qy	222 GGAAACTGCTCGGTGGCAGCGCGAGCCGCCACCGCGGACAGCTCGCTCCAGGCTGCGTCA 281
Dn	485 CGCGGCGGCGCGCGCGCGCGCGCGCGGGGGCGCGGGGAGCGCGCGCGCGCGCGGG 544
Qy	282 CGGACCTTCGCGGGTTCGGAGCGGAAAACGGCGACGGCTCAGATCCTTCAGTTCGATC 341
Dn	545 CGGGGCGGCGCGCGCGGGGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGG 604
Qy	342 GCCTCAGTGCGGTGTCCTCCCTTCGCTGGGAGATAGCGTTACGACGAGCGGCGAC 401
Dn	605 GGGCGCGCGGGGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 664
Qy	402 ACGGGGGCGGGGGCGGGGGCGGTTTCAGCCGATCCGCTCAGTATGACAGCGGTTGCGGGT 461



gene	translation="MLSGLSAHPDARVRAQVALNPNTPAEVLGLIAAQFFPREVLSNPG LPLRLARGLGSPAEQVTAALLPAPGAVVDSALRHEDYGVRTALAAAPGLSPE RYTALAGAGVQRAQRPDLPEFLRLQAEEDYVRKATRLPDLPGDLRLTLV TPDHTLVANVRDLDPDCLSLATGDPDVLATLARRDTLPASVREMLAAAEPL VRSALQAVTPESMLARETADPVRSAALRRPDLSPVELRHADSEGTVRITLL ENDLPDGLAALAAQEPDPLQVSAEELPAALEALLDESDVRLTILAVRDLG PERLAKADEASAVRGRHALAPTLPPELLRQLSADPNAGVRLITLRELPAAEVIRE LAADPDGVRLLVAGRAGLPAQLRQLAADPDENVRRVAG" /gene="DR1774" 1944. .2396 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1341.1" /db_xref="GI:6459561" /translation="MALGRQEKFLPLADOSTVINVQDPRGVSSREGIRPFEDFARF GASKGVESRDLHCDQSVSAVCHNVLLIGGADGFOCARLPMKAGOLSPRPHERSV RSDCTFRIHRWDQRHAGGSDGAGRIHRSAGSQTEQGRSYPTKSHA" /gene="DR1775" 2469. .4706 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="DNA helicase I1" /protein_id="AAFI1329.1" /db_xref="GI:6459549" /translation="MTSSAGPDLQALNPQAAADHTGPAIVLAGGKTRTLIY RIAHLGIVGHGEILLVFTPKAAEMREERAGHLVPGAGDLWMSTFSAQVRLIT YGHIGLRGFRVYIDDDQDDIIKEVMSIFGICAGTQPRVIRGLIDRAKSNLWPTDD LDRSEFFISGLPRDAAAEAYRYEVRRKQNAIDFGDLITETVRLDFKGVGLKQVQ NKAFTHVDYQDTRAQVELTSLASRDNLVVGDPDQSIYKFGADIQNLDFQK DYPDAKYMLHRYRSARVLEANKLIENNERDLTKLPVKEAQPVTFHRTDHR AEGDYADVTLRLHGRAMSEMAILYRNAQSRVIEESLRVQIPARIVGGVGYDR REIRDLIYARHATKATFAGLEAMSRADNVEPAALRFVMTSGYLDLROEGQBGQ AARONLDRGAHATKATFAGLEAMSRADNVEPAALRFVMTSGYLDLROEGQBGQ VRLNELSVSAEWSQDANVGGSIADPLDDAALLSSVDMDRKAENKGAPEDAVT LMLTLNAGLEFPVVFIVGVEQGLLESKGAISGPGSIEBERLFFVIGITRAMERLLM TAAQNRMQFQWNAEDSAFLDIEGLFDVDPYQPIETRAKTVKQIYPTPAATIA VKNTSPUTASLAVRGSGQVFKPKFGQGVLAAGVGERQEVTVHFASAGTKMLVKFA NLTKL" complement(4726. .5253) /gene="DR1776" complement(4726. .5253) /notes="similar to GP:3191993 percent identity: 45.21; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="MUT/mudix family protein" /protein_id="AAFI1330.1" /db_xref="GI:6459550" /translation="MPLGVGVWGEWGRVLPQMTDIRPLGGLKFSVRVALLCVNRGE LLVQTSANAPPEPFFLPFGALSTDEDAATCARREWEETGLLCGLRLVAVVNFEG PEKQKEHIGFYFLREELVPAELPHGAFSVLNDITGTPEWRLDELGARVYPLVRLD LSVPVGEVRHIVIRS" /gene="DR1777" 5309. .5608 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1335.1" /db_xref="GI:6459555" /translation="MVPVERQLVITALGLPORTLNGNDYLCQTEGEVVLWLPDPA	gene	translation="POLRVLRALTLPRRVELPCLPDRENAGRFWVPESARVVVVVDSAECAAPAHQCPA" 5810. .7105 /gene="DR1778" 5810. .7105 /notes="similar to GB:AB000657 percent identity: 70.88; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="3-isopropylmalate dehydratase, large subunit" /protein_id="AAFI1331.1" /db_xref="GI:6459551" /translation="MGMTIAEKMLAAHSGHDTVPQOLIECATDWWLCHETTPAALR MLEERGMDRVDPQKIVAVPDHSVPAMNIKAARMYQKLKSWQEKIEHFYDVGRI MLRVLENTGLMKFGQTLVSGDSTCNAGALGAFATGVSGLDLAGAIYAGKWFKVPET MLRVTKTNGPVTBKIDVLEIKOIGADGANVMVMWVGDIYDOLDMEGRFLTINNA IEAGKTCIVAVDDTTRAYMQRGVSPEQYTESQSDPDAKYKVVYIDAQVSPYTAQ PHIPSGRAGSDRIRVTHAYVSGSCINGRITDLRDVARILKGRKVAQDQVQMVVATQ AIKQAAQEGLEIFVDGASVSPSCAGLGMHSGVGLGPDICISSNNRNFVRMGD PSAQIVLASPATVAASAVEGFIADPRYDABGNDLNAAD" /gene="DR1779" 7112. .7453 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1336.1" /db_xref="GI:6459556" /translation="MNPADPPPSWLLTYFVIGWGLGLAVRDWKMALCGLTTLAVV FVLHGQPPVFRGLPGLMGLSALGLLALGVSRPLVLAARVGLAAVLTPLTWVVLNAG FLLPNVAQIS" /gene="DR1780" 7485. .8549 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1337.1" /db_xref="GI:6459557" /translation="MKGDMTLDPMKLLPPGVRPPEVTLHPGQGLDIRGLDEEAQAVQ REALKONPLASVREDTALVRGLIRGQALYVAVSETGDAEASWETRPFPDMPC VPGQDGNAPSTFALHYBEGRIDLEPDAPDLDLDAVEDAVREYHGTEQWABRHA GGPANELPARRVRLRLEETGLDAGLRVVALPETAAYLLHGGEMHYAPAPQNLAD APQSEEDDSFFDLFSLFDLHSVTALHLDHGRVEVGGEDLAPALRAALOSHLAGDL GAGDPKWEVALRGLFGLLEGDEYFSGLRVGLLRQMVADPFAALAAQAITPSAVTDDG TWTELDGDDGDEPDEGDE" /gene="DR1781" 8580. .9893 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1338.1" /db_xref="GI:6459558" /translation="MGEAKRKQGLMPTVFPFAELGRDGEVRVLOGPEDAGORALI EKALDSQSFGAAMDAAEYRTVSLGSRGGRFAIREDVERIPVPAALRQDGLGALGSA GOSQAVIPVEGVSRLREQRHSFEGENWQTLPPLRDPQVLMRALQQHPAFDIBGESL GQFADHWLEGRIDVTPDVGELDENGETLEFFTLVKEFHGQTPESWTAMHREMLEGQ	gene	translation="complement(4726. .5253) /gene="DR1776" complement(4726. .5253) /notes="similar to GP:3191993 percent identity: 45.21; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="MUT/mudix family protein" /protein_id="AAFI1330.1" /db_xref="GI:6459550" /translation="MPLGVGVWGEWGRVLPQMTDIRPLGGLKFSVRVALLCVNRGE LLVQTSANAPPEPFFLPFGALSTDEDAATCARREWEETGLLCGLRLVAVVNFEG PEKQKEHIGFYFLREELVPAELPHGAFSVLNDITGTPEWRLDELGARVYPLVRLD LSVPVGEVRHIVIRS" /gene="DR1777" 5309. .5608 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1335.1" /db_xref="GI:6459555" /translation="MVPVERQLVITALGLPORTLNGNDYLCQTEGEVVLWLPDPA	gene	translation="complement(4726. .5253) /gene="DR1776" complement(4726. .5253) /notes="similar to GP:3191993 percent identity: 45.21; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="MUT/mudix family protein" /protein_id="AAFI1330.1" /db_xref="GI:6459550" /translation="MPLGVGVWGEWGRVLPQMTDIRPLGGLKFSVRVALLCVNRGE LLVQTSANAPPEPFFLPFGALSTDEDAATCARREWEETGLLCGLRLVAVVNFEG PEKQKEHIGFYFLREELVPAELPHGAFSVLNDITGTPEWRLDELGARVYPLVRLD LSVPVGEVRHIVIRS" /gene="DR1777" 5309. .5608 /notes="identified by Glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAFI1335.1" /db_xref="GI:6459555" /translation="MVPVERQLVITALGLPORTLNGNDYLCQTEGEVVLWLPDPA
------	---	------	--	------	---	------	---

Query Match 10.1%; Score 52.4; DB 1; Length 12454;  
Best Local Similarity 48.6%; Pred. No. 71;  
Matches 143; Conservative 0; Mismatches 151; Indels 0; Gaps 0;  
Qy 23 GTGCGCCCGCGCGCCCATCACGAATCGCACTCCCGCGGGGTGACCACTGTCACCGAGTG 82  
Db 12011 GTGCGCCCGCGCGCGCGCATCAGGCCAGCCAGCCAGCAATCACGGCGCGAGGTC 11952



```

Qy 83 GGAGTTGAGCAGCAGCTCATCGCGCTGCGCGGATCCGGAACACCGGCGCAGGAT 142
Db 11951 GGAAGAGATGTCGCGAAGAGTTCTCGGTGAGATCACTGCTAGCGGTGCGGTCGGC 11892
Qy 143 CACGTGAGCAGCAGGATGCGCGCTGAGCTGAGCTGAGCGTGGCGGAGCTCGATCGTC 202
Db 11891 CACGATCAGCATCCCGACGCTGTCACGTACTCGTGTGTTGAGTGAATCAGCGGTAGTC 11832
Qy 203 GCCCGGCTCCACACACGAGGAGAACTGCGTGGTGGCAGCGGCGAGCCCGCGGACAG 262
Db 11831 CGCGTCACGCAACGCTGACGTCGCGCGGSCACAACTCGGACACTTCGAGCAGCTTCGC 11772
Qy 263 CTCGTCCAGGTGCGGTACGAGACTCTCGCGGTTCGGAGCGGAACGGCGAC 316
Db 11771 CTTGTCCAGCTGTGTCAGCGCGCCCTTCGCGTGTCTCGCGCGGCCCGGAGAGCGAC 11718

RESULT 40
AC143344
LOCUS Macaca mulatta clone CH250-268P23, DNA linear HTG 10-APR-2003
AC143344
ACCESSION AC143344
VERSION AC143344.1 Gi:29611668
KEYWORDS HTG; HTGS PHASE2; HTGS PGI.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
Csuors.M. and Milosavljevic,A.
1 (bases 1 to 152686)
Pooled genomic indexing (PGI): mathematical analysis and experiment
design
(in) Guigo, R. and Gusfield, D. (Eds.);
ALGORITHMS IN BIOFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI
2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;
Springer (2002)
2 (bases 1 to 152686)
Milosavljevic,A., Sodergren,E., Csuors,M., Li,B., Jackson,A.R.,
Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L.,
Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burch,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,I., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,X., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Egan,A., Earhart,C., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
Guaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homs,F.,
Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mathew,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokewo,S.,
Ogah,M., Okunuga,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,I., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Roife,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A.,

```

```

Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wang,O., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Wacington,S., Williams,G., Williamson,A.,
Wlezky,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorilla,S., Kucherlapati,R., Weinstein,G. and Gibbs,R.
Direct Submission
Unpublished
3 (bases 1 to 152686)
Worley,K.C.
Direct Submission
Submitted (08-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 152686)
Worley,K.C.
Direct Submission
Submitted (10-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: LDMN
Center clone name: CH250-268P23
----- Summary Statistics
Chemistry: Dye-primer Bodypy: Infinity% of reads
Chemistry: Dye-terminator Big Dye: Infinity% of reads
Consensus quality: 3754 bases at least Q40
Consensus quality: 4914 bases at least Q30
Consensus quality: 5932 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: The contigs are based on the application
* of the PGI method using the Human genome (NCBI build 31)
* as the comparative genome.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 152686: contig of 152686 bp in length.
FEATURES
source
1..152686
Location/Qualifiers
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/db_xref="taxon:9544"
/clone="CH250-268P23"
ORIGIN
Query Match 10.0%; Score 52.2; DB 2; Length 152686;
Best Local Similarity 44.2%; P: 0; Mismatches 239; Indels 0; Gaps 0;
Matches 189; Conservative
Qy 89 GAGCAGCAGCTCATCGCGCTGCGCGGATCCGGAACACCGGCGCAGGATCAGTCG 148
Db 56408 GAGCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 56467
Qy 149 CAGCAGCGCGGATCGCGCGCTTCGATCTCGACGGTTCGGCGCGGATCTCGTCGCGCGG 208
Db 56468 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 56527
Qy 209 CTCCACACACGAGGAGAACTCGGTCTCGTGTGACGCGCGCGCGCGCGCGCGCGCGTC 268

```





/notes="testis-specific; interacts with the DAZ protein,  
which is involved in male infertility"  
/codon\_start=1  
/product="quaking protein 3"  
/protein\_id="AAL36980.1"  
/db\_xref="GI:17224450"  
/translations="TSPRPPEPAAQAALSGLDPSRRRLVLRQPPAPART  
RPATRLPPRPRGRABFGRRRLGARRARASACGASARCPGPRASAREPE  
REPARGTTPPERPAAGRPRLSLGGGGGGGSELSELYGNGVGEAEAHPR  
LPDAADQERAEHQPAQLRLDLPPEAAAGRN"

## ORIGIN

Query Match 10.0%; Score 52; DB 9; Length 2489;  
Best Local Similarity 47.9%; Pred. No. 1.2e+02;  
Matches 181; Conservative 0; Mismatches 195; Indels 2; Gaps 1;  
  
QY 6 AGCAGCAACACCGCGTGTGTCGCCCGCGCGCCCATCACGAATGCCACTCCGCGCGGGTG 65  
DB 23 AGCCCGGGCACAGCGCGCGCGCGCTGAGCGGGCTCGACAGCGAGCGCGCGCG 82  
  
QY 66 ACCAGGTGACACAGTGGAGTTGACAGCAGCACTATGCCCGCTGCGCGCGCATGCCG 125  
DB 83 CCGCGCTCTCTCTGTCGCGCGCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 142  
  
QY 126 AAACACCGCGCGAGGATCAGTGCAGCAGCGCGATGCGCGCTCTGATCTCGACGTCGCG 185  
DB 143 AGACCGCTCTCCGCGCGCGCGCGCTGTGAGGAGCGCGCGCGCGCGCGCGAG 202  
  
QY 186 CGCAGCTGATCTGTCGCGCGCGCTCCACACAGGGGAACTGGCTCGTGGCAGCGCG 245  
DB 203 AGCGCGCGCGCTGGAGCGCG--TCGGGCGCGCGCGGAAAGTGCCTGCGGGGGCGGCG 260  
  
QY 246 AGCCCGACCGCGGACACTGTCAGCGTGGCTGTCAGGACCTCTCGCGCGTGGAGCG 305  
DB 261 GAGCGCGCTGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 320  
  
QY 306 GAAACGCGCGCGCTCAGATCCCTGTGTCGATCGCTCAGTGGCGGTGTCGCCCTT 365  
DB 321 GAGCGGACCGCGGTGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380  
  
QY 366 GCGCTGGGAGGATAGCG 383  
DB 381 CTCGCGCGCGCGCGCGCG 398

RESULT 43  
AL356119  
LOCUS  
DEFINITION Human DNA sequence from clone RP5-856H7 on chromosome 6, complete  
sequence.  
ACCESSION AL356119  
VERSION AL356119.31 GI:19351909  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 75152)  
Martin, S.  
Direct Submission  
Submitted (10-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Mar.11, 2002 this sequence version replaced gi:19309430.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP5-856H7 is from the library RPI-5 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2  
IMPORTANT: This sequence is not the entire insert of clone  
RP5-856H7. It may be shorter because we sequence overlapping  
sections only once, except for a short overlap.  
The true right end of clone RP5-856H7 is at 75152 in this sequence.  
The true left end of clone RPI-5LJ12 is at 74160 in this sequence.  
The true right end of clone RPI-495010 is at 2000 in this sequence.  
FEATURES  
Location/Qualifiers  
1..75152  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP5-856H7"  
/clone\_lib="RPI-5"  
23587..23649  
/note="Single clone region. Sequence from reads from a  
short insert library derived from a single PUC clone.  
Restriction digest data confirm the assembly."  
42971..43036  
/note="Single clone region. Sequence from reads from a  
short insert library derived from a single PUC clone.  
Sequence from overlapping clone RPI-164E20 (AL133542).  
Assembly confirmed by restriction digest."

## ORIGIN

Query Match 10.0%; Score 52; DB 9; Length 75152;  
Best Local Similarity 47.9%; Pred. No. 52;  
Matches 181; Conservative 0; Mismatches 195; Indels 2; Gaps 1;  
  
QY 6 AGCAGCAACACCGCGTGTGTCGCCCGCGCGCCCATCACGAATGCCACTCCGCGCGGGTG 65  
DB 23855 AGCCCGGGCACAGCGCGCGCGCGCTGAGCGGGCTCGACAGCGAGCGCGCGC 23914  
  
QY 66 ACCAGGTGACACAGTGGAGTTGAGCAGCGCGCTCATCGCGCGCTGCGCGCGCATGCCG 125  
DB 23915 CCGGCTCTCTCTGTCGCGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 23974  
  
QY 126 AAACACCGCGCGAGGATCAGCTGTCAGCAGCGCGATGCGCGCTCTCGATCTCGACGTCGCG 185  
DB 23975 AGACCGCGCTCCGCGCGCGCGCGCTCTGAGGAGCGCGCGCGCGCGCGCGAG 24034  
  
QY 186 CGCAGCTGATCTGTCGCGCGCGCTCCACACAGAGGGGAACTGGCTGCGTGGCAGCGCGC 245  
DB 24035 AGCGCGCGCGCTGGAGCGCG--TCGGGCGCGCGCGGAAAGTGCCTGCGCGGGCGCGCG 24092  
  
QY 246 AGCCCGACCGCGGACAGCTCTGTCAGGCTGCGCTCAGCGACTCTCGCGGGTCCGGAGCG 305  
DB 24093 GAGCGCGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24152  
  
QY 306 GAAACGCGCGCGCTCAGATCCCTGTGTCAGTGCATTCGGCTCAGTGGCGGTGTCGCCCTT 365  
DB 24153 GAGCGGAGCGCGGGTCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 24212  
  
QY 366 GCGCTGGGAGGATAGCG 383  
DB 24213 CTCGCGCGCGCGCGCGCG 24230

## RESULT 44

AC084804/c

## LOCUS

219952 bp DNA linear HTG 15-MAY-2002  
Mus musculus clone RP23-314K21 strain C57BL6/J, WORKING DRAFT  
SEQUENCE, 46 unordered pieces.

## DEFINITION

AC084804

## ACCESSION

AC084804.10 GI:17488575

## VERSION

HTG; HTGS PHASE1; HTGS DRAFT.

## KEYWORDS

Mus musculus (house mouse)

## SOURCE

Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## REFERENCE

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## AUTHORS

1 (bases 1 to 219952)

## TITLE

Grills, G., Han, J., Montgomery, K.T., Chiu, D., Decker, J., Fusina, M.,

## JOURNAL

Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,

## AUTHORS

Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

## TITLE

High-Throughput Mouse Sequencing

## JOURNAL

Unpublished

## AUTHORS

2 (bases 1 to 219952)

## TITLE

Grills, G., Han, J., Montgomery, K.T., Chiu, D., Decker, J., Fusina, M.,

## JOURNAL

Goltz, J., Haider, A., Hall, L., Ioshikhes, I.P., Lee, E., Long, J.,

## AUTHORS

Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.

## TITLE

Direct Submission

## JOURNAL

Submitted (18-NOV-2000) Department of Molecular Genetics, Albert

## AUTHORS

Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,

## TITLE

Bronx, NY 10461, USA

## JOURNAL

On Dec 11, 2001 this sequence version replaced gi:13560369.

## COMMENT

-----Genome Center

## AUTHORS

Center: Harvard Partners Genome Center

## TITLE

Center Code: HPGC

## JOURNAL

Web site: <http://www.hpcg.org/Sequence/mouse.html>

## AUTHORS

Contact: [hpgc@mcg.harvard.edu](mailto:hpgc@mcg.harvard.edu)

## TITLE

-----Summary Statistics

## JOURNAL

Center project name: AEZ

## AUTHORS

Sequencing vector: pUC18; L08752

## TITLE

Chemistry: Dye-terminator Big Dye; 100%

## JOURNAL

\*Consensus quality: 207914 at least Q20

## AUTHORS

\*Consensus quality: 205077 at least Q30

## TITLE

\*Consensus quality: 200559 at least Q40

## JOURNAL

\*Estimated insert size: agarose-FP - N/A

## AUTHORS

\*\*Estimated insert size: 219052 - sum-of-contigs

## TITLE

Quality coverage: agarose-FP - N/A

## JOURNAL

Quality coverage: 10.7 x in Q20 bases; sum-of-contigs estimation

## AUTHORS

-----

## TITLE

\* NOTE: This is a 'working draft' sequence. It currently

## JOURNAL

\* consists of 46 contigs. The true order of the pieces

## AUTHORS

\* is not known and their order in this sequence record is

## TITLE

\* arbitrary. Gaps between the contigs are represented as

## JOURNAL

\* runs of N, but the exact sizes of the gaps are unknown.

## AUTHORS

\* This record will be updated with the finished sequence

## TITLE

\* as soon as it is available and the accession number will

## JOURNAL

\* be preserved.

## AUTHORS

\* 1 26658: contig of 26658 bp in length

## TITLE

\* 26659 26678: gap of unknown length

## JOURNAL

\* 26679 53221: contig of 26643 bp in length

## AUTHORS

\* 53322 53341: gap of unknown length

## TITLE

\* 53342 75821: contig of 24480 bp in length

## JOURNAL

\* 75822 75841: gap of unknown length

## AUTHORS

\* 75842 99596: contig of 23755 bp in length

## TITLE

\* 99597 99616: gap of unknown length

## JOURNAL

\* 99617 120496: contig of 20880 bp in length

## AUTHORS

\* 120497 120516: gap of unknown length

## TITLE

\* 120517 132931: contig of 12415 bp in length

## JOURNAL

\* 132932 132952: gap of unknown length

## AUTHORS

\* 132953 144027: contig of 11076 bp in length

## TITLE

\* 144028 144047: gap of unknown length

## JOURNAL

\* 144048 153071: contig of 9024 bp in length

## AUTHORS

\* 153072 153091: gap of unknown length

## TITLE

\* 153092 162146: contig of 9055 bp in length

## JOURNAL

\* 162147 162166: gap of unknown length

## AUTHORS

\* 162167 167980: contig of 5814 bp in length

## TITLE

\* 167980 219680: contig of 273 bp in length.

## JOURNAL

\* 219680 219952: contig of 273 bp in length.

## AUTHORS

\* 219952 219952: contig of 273 bp in length.

## TITLE

\* 219952 219952: contig of 273 bp in length.

## JOURNAL

\* 219952 219952: contig of 273 bp in length.

## AUTHORS

\* 219952 219952: contig of 273 bp in length.

## TITLE

\* 219952 219952: contig of 273 bp in length.

## JOURNAL

\* 219952 219952: contig of 273 bp in length.

## AUTHORS

\* 219952 219952: contig of 273 bp in length.

## TITLE

\* 219952 219952: contig of 273 bp in length.

## JOURNAL

\* 219952 219952: contig of 273 bp in length.

## AUTHORS

\* 219952 219952: contig of 273 bp in length.

## TITLE

\* 219952 219952: contig of 273 bp in length.

## JOURNAL

\* 219952 219952: contig of 273 bp in length.

## AUTHORS

\* 219952 219952: contig of 273 bp in length.

## TITLE

\* 219952 219952: contig of 273 bp in length.

## JOURNAL

\* 219952 219952: contig of 273 bp in length.

## AUTHORS

\* 219952 219952: contig of 273 bp in length.

## TITLE

\* 219952 219952: contig of 273 bp in length.

## JOURNAL

\* 219952 219952: contig of 273 bp in length.

FEATURES

Location/Qualifiers

```

source
1. 219952
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/clone="RP23-314K21"
/sex="male"
1. .26658
/misc_feature
/notes="assembly_name:Contig192"
26679. .53321
/misc_feature
/notes="assembly_name:Contig191"
53342. .75821
/misc_feature
/notes="assembly_name:Contig190"
75842. .99596
/misc_feature
/notes="assembly_name:Contig189"
99617. .120496
/misc_feature
/notes="assembly_name:Contig188"
120517. .132931
/misc_feature
/notes="assembly_name:Contig187"
132952. .144027
/misc_feature
/notes="assembly_name:Contig186"
144048. .153071
/misc_feature
/notes="assembly_name:Contig185"
153092. .162146
/misc_feature
/notes="assembly_name:Contig184"
162167. .167980
/misc_feature
/notes="assembly_name:Contig183"
168001. .173254
/misc_feature
/notes="assembly_name:Contig182"
173275. .179246
/misc_feature
/notes="assembly_name:Contig181"
179267. .180707
/misc_feature
/notes="assembly_name:Contig180"
180728. .184387
/misc_feature
/notes="assembly_name:Contig179"
184408. .189237
/misc_feature
/notes="assembly_name:Contig178"
189258. .192394
/misc_feature
/notes="assembly_name:Contig177
clone_end:SP6
vector_side:left"
193015. .197265
/misc_feature
/notes="assembly_name:Contig176"
197286. .197589
/misc_feature
/notes="assembly_name:Contig175"
197610. .199581
/misc_feature
/notes="assembly_name:Contig174"
199602. .201006
/misc_feature
/notes="assembly_name:Contig173"
201027. .202394
/misc_feature
/notes="assembly_name:Contig172
clone_end:T7
vector_side:right"
202415. .203415
/misc_feature
/notes="assembly_name:Contig171"
203436. .204824
/misc_feature
/notes="assembly_name:Contig170"
204845. .205380

Query Match 10.0%; Score 52; DB 2; Length 219952;
Best Local Similarity 47.6%; Pred.No. 40;
Matches 201; Conservative 0; Mismatches 218; Indels 3; Gaps 2;

QY 56 CGCGGGGTGACCAAGTCTGACACAGTGGAGTGTGAGCAGCAGCTCATCGCGCCTGTGCG 115
Db 189210 GTTGGCGGGGCGGGGCGGGGCGCGCGCGGGGCGCCCCCGTGGCGCGCGCG 189151

QY 116 CGGATGCGGAAACACCGGGCGAGATACGTGACGACCGGCATCGCGCCTCGATCTC 175
Db 189150 GCGCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 189092

QY 176 GACGTCGCGCGCAGCTCATCTCGTGG-.CCCGGCTCCACACCGAGGGGAATGGCTC 233

```





```

QY 304 CGGAACGCGCAGCGGTCTAG--ATCCCTGTCACTCCATCGGCTCAGTCCGCGGTCTGCC 361
Db 4476 TGAACGCGGTTCTGTCACGCCAGCGCGCCATCCGGGTCTTGTCCCGGAGATCGACCC 4417
QY 362 CTTGCGCTGGGAGGATAGCGGTTCACGACGAGCGGCACACGCGGGCGGGCGGGGG 421
Db 4416 GCGGTTCCGGGTCTGACGACGCGGTTCGACGCGGCGGACGCGGCACCGGTGCGCGCGGTC 4357
QY 422 CGGTTTCAGCCGATCCGCTCGATACGACGCGGTTCGCGGCGGTTCGCGGAGATCC 481
Db 4356 AGGTCGGTGCAGCTGCTGATGACCTCGATTCGCGACTCCGGGAGCGCGGTGATGCC 4297
QY 482 GTACCGCGCGGACCGCTCGGCGGCGCGCGCG 514
Db 4296 GACGCGCGGCGCGCGCGCGCGCGCGCGCGAG 4264

RESULT 47
AC010821 208936 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-3K18, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC010821
ACCESSION AC010821.4 GI:9120942
VERSION HTG; HTGS PHASE0.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 208936)
Birren,B., Linton,L., Nusbaum,C. and Lander,B.
Homo sapiens, clone RP11-3K18
Unpublished
2 (bases 1 to 208936)
Birren,B., Linton,L., Nusbaum,C., Lander,B., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Becker,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melarim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul.13, 2000 this sequence version replaced gi:8225265.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTER
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2764
Center clone name: 3_K_18
-----
* NOTE: This record contains 229 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

```

```

* be preserved.
1
697: contig of 697 bp in length
797: gap of 100 bp
1497: contig of 700 bp in length
1597: gap of 100 bp
2265: contig of 668 bp in length
2365: gap of 100 bp
3044: contig of 679 bp in length
3144: gap of 100 bp
3343: contig of 695 bp in length
3939: gap of 100 bp
4642: contig of 703 bp in length
4742: gap of 100 bp
5442: contig of 704 bp in length
5546: gap of 100 bp
6268: contig of 722 bp in length
6368: gap of 100 bp
7066: contig of 698 bp in length
7166: gap of 100 bp
7856: contig of 690 bp in length
7956: gap of 100 bp
8640: contig of 684 bp in length
8740: gap of 100 bp
9415: contig of 675 bp in length
9515: gap of 100 bp
10181: contig of 666 bp in length
10281: gap of 100 bp
10969: contig of 688 bp in length
11069: gap of 100 bp
11760: contig of 691 bp in length
11860: gap of 100 bp
12565: contig of 705 bp in length
12665: gap of 100 bp
13372: contig of 707 bp in length
13472: gap of 100 bp
14162: contig of 690 bp in length
14262: gap of 100 bp
14949: contig of 687 bp in length
15049: gap of 100 bp
15750: contig of 701 bp in length
15850: gap of 100 bp
16524: contig of 674 bp in length
16624: gap of 100 bp
17323: contig of 705 bp in length
17423: gap of 100 bp
18125: contig of 696 bp in length
18225: gap of 100 bp
18911: contig of 686 bp in length
19011: gap of 100 bp
19698: contig of 687 bp in length
19798: gap of 100 bp
20498: contig of 700 bp in length
20598: gap of 100 bp
21313: contig of 715 bp in length
21413: gap of 100 bp
22130: contig of 717 bp in length
22230: gap of 100 bp
22917: contig of 687 bp in length
23017: gap of 100 bp
23713: contig of 696 bp in length
23813: gap of 100 bp
24511: contig of 698 bp in length
24611: gap of 100 bp
25296: contig of 685 bp in length
25396: gap of 100 bp
26075: contig of 678 bp in length
26174: gap of 100 bp
26796: contig of 622 bp in length
26896: gap of 100 bp
27596: contig of 700 bp in length
27696: gap of 100 bp
28400: contig of 704 bp in length
28500: gap of 100 bp

```



28501 29220: contig of 720 bp in length  
29221 29320: gap of 100 bp  
29321 30032: contig of 712 bp in length  
30033 30132: gap of 100 bp  
30133 30825: contig of 693 bp in length  
30826 30925: gap of 100 bp  
30926 31627: contig of 702 bp in length  
31628 31727: gap of 100 bp  
31728 32408: contig of 681 bp in length  
32409 32508: gap of 100 bp  
32509 33192: contig of 684 bp in length  
33193 33292: gap of 100 bp  
33293 33968: contig of 676 bp in length  
33969 34068: gap of 100 bp  
34069 34744: contig of 676 bp in length  
34745 34844: gap of 100 bp  
34845 35226: contig of 682 bp in length  
35227 35626: gap of 100 bp  
35627 36322: contig of 696 bp in length  
36323 36422: gap of 100 bp  
36423 37118: contig of 696 bp in length  
37119 37218: gap of 100 bp  
37219 37946: contig of 728 bp in length  
37947 38046: gap of 100 bp  
38047 38736: contig of 690 bp in length  
38737 38836: gap of 100 bp  
38837 39531: contig of 695 bp in length  
39532 39631: gap of 100 bp  
39632 40118: contig of 687 bp in length  
40119 40418: gap of 100 bp  
40419 41109: contig of 691 bp in length  
41110 41209: gap of 100 bp  
41210 41900: contig of 691 bp in length  
41901 42000: gap of 100 bp  
42001 42619: contig of 619 bp in length  
42620 42719: gap of 100 bp  
42720 43400: contig of 681 bp in length  
43401 43500: gap of 100 bp  
43501 44116: contig of 616 bp in length  
44117 44216: gap of 100 bp  
44217 44916: contig of 700 bp in length  
44917 45016: gap of 100 bp  
45017 45713: contig of 697 bp in length  
45714 45813: gap of 100 bp  
45814 46497: contig of 684 bp in length  
46498 46597: gap of 100 bp  
46598 47293: contig of 696 bp in length  
47294 47393: gap of 100 bp  
47394 48098: contig of 705 bp in length  
48099 48198: gap of 100 bp  
48199 48892: contig of 694 bp in length  
48893 48992: gap of 100 bp  
48993 49666: contig of 674 bp in length  
49667 49766: gap of 100 bp  
49767 50453: contig of 687 bp in length  
50454 51229: gap of 100 bp  
51230 51329: contig of 676 bp in length  
51330 52017: contig of 688 bp in length  
52018 52117: gap of 100 bp  
52118 52803: contig of 686 bp in length  
52804 52903: gap of 100 bp  
52904 53610: contig of 707 bp in length  
53611 53710: gap of 100 bp  
53711 54406: contig of 696 bp in length  
54407 54506: gap of 100 bp  
54507 55214: contig of 708 bp in length  
55215 55314: gap of 100 bp  
55315 56029: contig of 715 bp in length  
56030 56129: gap of 100 bp  
56130 56834: contig of 705 bp in length  
56835 56934: gap of 100 bp  
56935 57626: contig of 692 bp in length

Query Match 10.0%; Score 51.8; DB 2; Length 208936;  
Best Local Similarity 47.0%; Pred. No. 43;  
Matches 151; Conservative 0; Mismatches 169; Indels 1; Gaps 1;  
  
QY 197 CTGCTCGCCCGGTCCTCCACACACAGGGAACTGGCTCGTGGCAGCGGCGAGCCCGCCGCG 256  
Db 100266 CCCCCCGCGGCCCGCCCGNCCNCCGNGGCGCGGCGCGCGCGCGCGCGCGCGCGCG 100325  
QY 257 GGACAGCTCGTCCAGGCTCGCTCAGCGACCTCTCTCGCGGCTCGGAGCGGAAACCGGCAC 316  
Db 100326 GNGGCGCCNCGGGCG 100385  
QY 317 GGCTCAGATCCCTGTCTAGTCGATCGGCTCACTGCGGTCGTCCTCCCTTGGCTGGGAGG 376  
Db 100386 CCCCCCGCGCGGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCGCGCGCG 100445  
QY 377 ATAGCGGTTTCAGACAGAGCGGACACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 435  
Db 100446 CCGGCG 100505  
QY 436 CGCTCGATGACACAGCGGCTGCGGGGTGCGGGGTCGGCGAGATCGGTACCGCGCGGAC 495  
Db 100506 CCGGCG 100565  
QY 496 GCCTCG 516  
Db 100566 GNGGCG 100586

## RESULT 48

AL646079  
LOCUS Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;  
DEFINITION segment 4/11.  
ACCESSION AL646079 AL646053  
VERSION AL646079.1 GI:17430956  
KEYWORDS  
SOURCE Ralstonia solanacearum  
ORGANISM Ralstonia solanacearum  
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
Burkholderiaceae; Ralstonia.  
REFERENCE 1  
AUTHORS Salanoubat, M., Genin, S., Artiguenave, P., Gouzy, J., Mangenot, S.,  
Arlat, M., Billault, A., Brottier, P., Camus, J.-C., Cattoilco, L.,  
Chandler, M., Choise, N., Claudel-Renard, C., Cunnac, S., Demange, N.,  
Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T.,  
Sigulier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M.,  
Weissenbach, J. and Boucher, C.A.  
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum  
NATURE 415 (6871), 497-502 (2002)  
MEDLINE 21681879  
PUBMED 11823852  
REFERENCE 2 (bases 1 to 213050)  
AUTHORS Boucher, C.A.  
TITLE Direct Submission  
JOURNAL

## COMMENT

Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston  
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie  
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,  
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean  
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS  
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA  
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,  
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,  
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire  
INRA, BP27, F31326 Castanet-Tolosan Cedex  
Christian.Boucher@toulouse.inra.fr  
http://sequence.toulouse.inra.fr/R.solanacearum.html.

## FEATURES

source  
1. .213050  
Location/Qualifiers  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GMI1000"

```

/db_xref="taxon:305"
/plasmid="megaplasmid"
100..1632
/gene="RSp0487"
/notes="synonym: RS00365"
100..1632
/gene="RSp0487"
/function="macromolecule metabolism; macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="HYPOTHETICAL TRANSMEMBRANE PROTEIN"
/protein_id="CADI7638.1"
/db_xref="GI:17430957"
/db_xref="GOA:Q8XS17"
/db_xref="SPTRMBL:Q8XS17"
/translations="MDAPSGRAVIRRAVGRLLLVVGLAVLCIAFNWRHDSR
QLDVPAPLAVDGAARALGEARTISSTDAQNADQFRLHMLAARYPKAH
AMQRPVGFSLIYTKGSDPSLKPILLMAHQDVVPVAPGTEGVPPFAGVVKOG
VMVGRGMDKGNLIAQMAAELLAAAGFRPRTIFAFGADEEVGGERGAIAALL
KSRGEOLAVIDRGLITGVVPGLTKPAALIGVTEKGLSVALKLSATPGHSSMPPA
PGSATAWMSAALKHLDLQDLQFAGIGVTEMPETIAPKRGINRVALNSNLWLFPELV
RKLQGLPSTNVAITTLTIYVQGNKONVUPGRABATVFNRLPFGDSSVIAHVE
QAVRSVAGHFELALPGVSEAAVPSPTOSAYQLIGRTVEVPGTVVAPGLMVG
TSRHHVIGISDHVFSVRPREDIARFHTGNTERISEANVELIRFYHRLIHQAADA
GA"
complement (1751..3148)
/gene="czcS"
/notes="synonyms: RS00366, RSp0488"
complement (1751..3148)
/gene="czcS"
/EC number="2.7.3.-"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE"
/transcription="REGULATOR PROTEIN"
/protein_id="CADI7639.1"
/db_xref="GI:17430958"
/db_xref="GOA:Q8XS16"
/db_xref="SPTRMBL:Q8XS16"
/translations="MKRWPSLAARLVAFVLVAGVMLGAGTYLYALNAQLRATDD
AMLLGIVOLRHCADKTQSDIDVDRARYTEVVSCHGLILRVIPGGRATVEANPE
GEVAPAAVPPVYLASEALHTWNSGTGIASRAIAAQAALGNHRTVTVLVAQGDG
RAALMASHRIHFLAVIAGVAALLSLMLVRSITLKLARIQRQASVVSQDLTRLT
LEGAPRSLDLAANMLTRLQDGFARLSQFSADLAHDFRTPLANLIGQETVLAHP
RPAEYRVALESLEEYARLSRMIEDMLFLARADHARVALSVAFPLDARHEIOVLVDYF
EVAADRDTIQVCGNGRVADITLLRRAINLLDNLARHTPGGARIDVAOVLPFST
AIAVRNPGPGIIPAASIPLLPDRFYRGDPARANSERSAGLGLAIVQITIMDLHG3TAQAS
SPPGGLTEPLIFPA"
complement (3145..3819)
/gene="czcB"
/notes="synonyms: RS00367, RSp0489"
complement (3145..3819)
/gene="czcB"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage

```

```

predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CADI7640.1"
/db_xref="GI:17430959"
/db_xref="GOA:Q8XS15"
/db_xref="SPTRMBL:Q8XS15"
/translations="MRLLIVEDRPKAGDYLLKGLTSGFVADLARSQDGLYOAMEHD
YDLVLDVMLPGMDGQVIBELRRKTKPVLFETARDELSDELKGLGLGADDYLVKPP
AFAEVLAVRHITLARGPMRESEVLEADLIDITIKRVTAGRRIDLTAKAYALLHL
ARRTGEVLSRSLISSQVWVNFDSNTNVDAIRLRARVDVDFEHLIHLTKMGIV
LDTVAR"
4089..4394
/gene="RSp0490"
/notes="synonym: RS00368"
4089..4394
/gene="RSp0490"
/function="miscellaneous; hypothetical/global homology"
/notes="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
/protein_id="CADI7641.1"
/db_xref="GI:17430960"
/db_xref="GOA:Q8XS14"
/db_xref="SPTRMBL:Q8XS14"
/translations="MSVVRNCLPLAARAAFLACGATATAVSQFALSPLQVHCARP
YTDGARGTRPYTDGGAQAEATLQASVGPDRDPDSCGMSRSPDRDRDPTDGAQA"
4400..5725
/gene="czcC"
/notes="synonyms: RS00369, RSp0491"
4400..5725
/gene="czcC"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
RESISTANCE PROTEIN"
/protein_id="CADI7642.1"
/db_xref="GI:17430961"
/db_xref="GOA:Q8XS13"
/db_xref="SPTRMBL:Q8XS13"
/translations="MWTHPCIRAVALS CALLVRSGLAARAPAGAAAASDTAAPAQS
APASVPGAHERTLQQAIDRALATHPALAAGYAVRAGSAGVDAQSRLRNPAAFLT
EDVRDRATTAQNLITFLELGGKRAARTRAAELARDAGQDAAVLAALRADVYTRAF
ELSTAQRLRLAQSDALAAQAAAVRVYQAGKVSPEQTKAAVARAGAAIELRDAQ
ADLVAQRLSAFWGGA VAEADGSAETLPVFPPEAHRYLAATAAEPAVALASTPRA
LRARLETHQALVNVRSKRVDPVTLVGAKDSDSGANANMVVGIAPVLPDRNOG
SLLEAQRLAADAEDYALAEQAALSQDVARLEAQAQVAGLRQDVLPGAGRAYDA
ARIGFEAGKFNFLDLDQAQRTLFQARAQYLAALSRAHQAAAGIDRIILGR"
5752..6966
/gene="czcB"
/notes="synonyms: RS00370, RSp0492"
5752..6966
/gene="czcB"
/function="cell processes; protection responses"
/notes="Product confidence : probable
Gene name confidence : probable
Predicted by Codon usage

```



```
SWALL:AA181961 (EMBL:AE010279) (236 aa) fasta scores: E():
4e-15, 33.04% id in 233 aa
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE40070.1"
/db_xref="GI:3356543"
/translation="MSTALISWAGAGTKLDEIDLLLOQWIDAGROVHEADSKRLTAA
GFFVPEGRAGPAAVKLCAKSLHKSSELGLVALDVPDVERARREIQSRRAQVAG
GEVLVSVMGVDMFVWCRTDHTTFFGVVVVGGVIGAGVPEIRMAFLPSARDAE
RALRHKAFFIIDGARGEPADIGAFARLIADVSEFYRRSHLHIAELDLPVVMVRGRH
LDPGMVADASIIILQKTF"
complement(3388..4344)
/locus_tag="BPP0662"
/complement(3388..4344)
/locus_tag="BPP0662"
/notes="Similar to Escherichia coli and Escherichia coli
O157:H7 glycine cleavage system transcriptional activator
Gcva or B2808 or Z4125 or Ec83668 SWALL:Gcva ECOLI
(SWALL:P32064) (305 aa) fasta scores: E(): 4e-27, 32.33%
id in 300 aa, and to Rhizobium loti glycine cleavage
system transcription activator Mlr1973 SWALL:Q98J99
(EMBL:AP002399) (303 aa) fasta scores: E(): 2.2e-28,
36.33% id in 300 aa"
/codon_start=1
/transl_table=11
/product="LysR-family transcriptional regulator"
/protein_id="CAE40071.1"
/db_xref="GI:33565438"
/translation="MQRPPNALRSPESAPRHESFLKAAELHVTGAVSLVKSLE
SYGLFELMRSHRGVAITPGKAPAEIDLEAFNKIGIATDRVQHPGDRGLSICHPIT
FAAHLTPWQVQAALSTQINVTTLAPELESTDDYFVKLGDPDPAASQACGGV
AAERVLVDFTFICSAFLQHGELGDDIERFLIHAOLRPQDMNWLQSVQGPAL
ROAPALVPESLITAYNAAGSAGMAIGIHAFIADDIASGRIVAPFPQVRKSHMGFNII
FNAQRAARVPRTEQIRVMVQERNKTRATDTTAAALRKSAFAE"
complement(3445..4083)
/locus_tag="BPP0662"
/notes="HMMpfam hit to PF03466, LysR substrate binding
domain, score 3.7e-16"
complement(4147..4326)
/locus_tag="BPP0662"
/notes="HMMpfam hit to PF00126, Bacterial regulatory
helix-turn-helix protein, lysR family, score 6e-23"
complement(4192..4284)
/locus_tag="BPP0662"
/notes="PS00044 Bacterial regulatory proteins, lysR family
signature."
complement(4564..5421)
/locus_tag="BPP0663"
complement(4564..5421)
/locus_tag="BPP0663"
/notes="Similar to Ralstonia solanacearum putative
hydrolyase protein Rsp0790 or R501917 SWALL:Q8ARPI
(EMBL:AL646080) (305 aa) fasta scores: E(): 2.3e-16,
33.61% id in 238 aa, and to Streptomyces coelicolor
putative hydrolase Sc04395 or Scd10.27 SWALL:Q9K3V0
(EMBL:AL359988) (352 aa) fasta scores: E(): 4.5e-15, 31%
id in 258 aa"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="CAE40072.1"
/db_xref="GI:33565439"
/translation="MQRDYLRLAHGLDFHWTSQGLPVVMLHGIRGYAETPAGI
AAALQPEVRVIAFDQGRGTDWDADCNVYTDYVADLAADQLSLARFDLHSGMG
GINAVYAAARHSGVGLVEDPAGFSDGCAIRIRRELAETPAGFAGVDAASDFM
RSNPSVSEARQRLSMLKPAPEGVYWRVYDHAGIATRLHPDPARVVDIAAYVAR
IGCTLLRLGRSDYIQPAMVERNALNPRIEATADAGHYIHDDQPAAPARHVCRF
LKPAPAGA"
complement(4591..5268)
/locus_tag="BPP0663"
/notes="HMMpfam hit to PF00561, alpha/beta hydrolase fold,
score 1e-25"

gene
complement(5447..6049)
/locus_tag="BPP0664"
complement(5447..6049)
/locus_tag="BPP0664"
/notes="Similar to Rhizobium loti hypothetical protein
Mlr6789 SWALL:Q988D2 (EMBL:AP003010) (196 aa) fasta
scores: E(): 3e-22, 38.71% id in 186 aa, and to
Agrobacterium tumefaciens hypothetical protein Atu4237 or
Aqr 1.1248 SWALL:AAK89202 (EMBL:AE009353) (200 aa) fasta
scores: E(): 2.1e-14, 39.08% id in 197 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAE40073.1"
/db_xref="GI:33565440"
/translation="MKTNFSVYHVRKMYLQVEDTLAGTGAPADGEPILRKIVLAACI
HNPYKGYVLDLTLWIDASPGIGABRRVRAAQQAISYKCALVADGDEYEHGN
ALTNPAAVNPDRADAGAKSWPSTGKRGPGTIIIDIPLAHKDALVVRSHYDVSQAQ
NDGPNDEVLVWAFATRRLNARLGLAACDVQHDGLN"
complement(5582..5611)
/locus_tag="BPP0664"
/notes="PS00758 ArgE / dapE / ACV1 / CPQ2 / yscS family"

misc_feature
Query Match 10.0%; Score 51.8; DB 1; Length 348257;
Best Local Similarity 48.5%; Pred. No. 38;
Matches 143; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 112 GCGCGGCATGCCGAACACCGCGCCAGGATCAGTGCAGCAGCGGATCGCGCCTCGA 171
DB 79205 GCGCGGCCTGGACGAGAAGCGCTGGCAGCGCTGGCGGCAAGCATCGATCGCGCGGTGC 79264

QY 172 TCTCGACGGTGGCGCGCAGCTCGATCTCGTCGCGCGGCTCCACACACAGGGGAACTGGC 231
DB 79265 GCTCGATATCGCGATCGTCCGCGAGCCTTGGCGGGGTTCCGCGGCTGTGATCGCGC 79324

QY 232 TCGGTGCGACGCGCGCCCGCGGACAGCTCGTCCAGGCTGGGCTACGACCTCTC 291
DB 79325 CGACAGCAGCGCGCGCGCGCGCGATCGCGCGCTCGCGCACGCTCCGCCACGGT 79384

QY 292 GCGGGTGGGAGCGGAAACCGCACGCGTCAGATCCTGTCTGTCATCGCATCGGCTCAGTGC 351
DB 79385 GCGCGTGGCGAGGACACCGCGCGCGCGCTTCCCGGCGCGCGCGCGCGGTCC 79444

QY 352 CGGTGCTCCCTTTGGCTGGGAGGATAGCGGTTTCAGACGAGCGGCGCACACCGC 406
DB 79445 GCGCGGACGCTCGCGCTGGCGGTGTGCTGTCTCGCGGTGTGCGCGCTGCTC 79499

RESULT 50
AX805262/c
LOCUS
AX805262 1293 bp DNA linear PAT 25-NOV-2003
DEFINITION
Sequence 17 from Patent WO03060128.
ACCESSION
AX805262
VERSION
AX805262.1 GI:38522381
KEYWORDS
uncultured actinomycete
uncultured actinomycete
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
environmental samples.
ORGANISM
1
REFERENCE
1
AUTHORS
Farnet, C.M., Staffa, A. and Zazopoulos, E.
TITLE
Compositions, methods and systems for discovery of lipopeptides
JOURNAL
Patent: WO 03060128-A 17 JUL-2003;
Ecopia Biosciences Inc. (CA)
location/Qualifiers
FEATURES
source
1..1293
/organism="uncultured actinomycete"
/mol_type="unassigned DNA"
/db_xref="taxon:100235"

ORIGIN
Query Match 9.9%; Score 51.6; DB 6; Length 1293;
Best Local Similarity 48.0%; Pred. No. 1.6e+02;
```

	Matches	212;	Conservative	0;	Mismatches	224;	Indels	6;	Gaps	2;
Qy	23	GTGCCCCCGCCG	CCATCAGCACTG	CCACTCGCGCGGGT	GACCAAGTTCGAC	CAAGTG	82			
Db	522	GTGCGCTCTGTCC	AGCTCGAGACCC	CGTCCGCGGGT	CCGCGCGCGCGG	CGC	463			
Qy	83	GGAGTTGAGCAGC	AGCTCATCGCGCT	CGCGCGCATCG	CGAAACACCGG	CCAGGAT	142			
Db	462	GGCGTAGACCTG	CGCCACCCCGG	CGCGGTAGACG	GAAGCGCTCC	AGCACGGC	403			
Qy	143	CACGTGAGCAGC	AGCGATGCGCGCT	CGATCTGACG	GTGCGCGAGCT	CGATCTCGTC	202			
Db	402	GTGCTGTGTAG	CGCGTGAACCA	CGCGCGGTCC	CGCGCGAGCAG	CGCGCTGGCG	343			
Qy	203	GCCCGGCTCCCA	CACAGCGGAACT	CGCTCGGTGG	CAGCGCGCGCG	CGCGACAG	262			
Db	342	GCTCAGCGGG	CGCCCGCGG	CGCGACCTCG	CGCCCATCG	CGCGCGCG	283			
Qy	263	CTCGTCCAGG	CTGGCTCAGG	ACTCTTCGCG	GTGCGGAGCG	GAAACGCGC	319			
Db	282	GCGCGCGCGCT	CCGGTGGGCT	CCGCGCGAG	GTCAACGACT	CCAGCGGCTC	223			
Qy	320	TCAGATCCCTG	TCAGTCGATCG	GTGCTCAGT	CGCGGTGCT	CGCCCTTGGG	379			
Db	222	GGCGGCGCG	CGAGCTGCCA	CACTGCTCG	ATCGTCTCG	AGCGCGCAG	163			
Qy	380	GCGGTTACAG	ACGACGCGCG	CGCGCGCGCG	CGCGCGCGCG	CGGTTACGCG	436			
Db	162	CTCGGTCTCG	GCACGCTGTC	CGGACCGCG	CGGTTCGAG	CGCGCGCG	103			
Qy	437	GCTCGATGAC	CACAGCGGCT	CGCG			458			
Db	102	GATCTCGTAG	CAGCGCGAG				81			

Search completed: June 27, 2004, 19:39:18  
Job time : 2470.78 secs



```
FEATURE:
; NAME/KEY: CDS
; LOCATION: 190..1398
US-08-812-412-1

Query Match      30.0%; Score 149.8; DB 2; Length 1401;
Best Local Similarity 65.3%; Pred. No. 1.2e-22;
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GGATCCACTGTCGCCAGCGCGGCGGAACTCTCTGCTCTTCCGCAATGATCTCTCGCGT 223
DB 1390 GGATCCAGCGCGCCCTCTCTGCTGAACTCCGCTCTCGGGGCTGGACTTCTGTAGGT 1331
QY 224 GGTTCACGCGAAGAGCAGCGCTAGTCCACCGGTGCGGCGTGAACGCGCTCCGGGTGC 283
DB 1330 GGTTCACGCGAAGAGCAGCGCTAGTCCGCGGTAGGCGGCCCGGAACTCTCGCGGAGC 1271
QY 284 GCACCGGATGTGTGTCGCGGGGTGAGCGCGCCCTCTTGGCGGGGTGCTTCGCACA 343
DB 1270 GGATCGGGATGTGGGACCCCGGGAGCCGACCGCGGATCTTGGCGGCTGTGTCTAGTA 1211
QY 344 CCCAGGAGACGAGTCCGAGCCGATGCCGAGAAAGTTCTGTCACGCTGGCGCTCTTCGCCG 403
DB 1210 CGCAGCGCAGAGATCGGGGCCGATCCCGCATAGTTGACACTGTGGCGCTCTTGGCGG 1151
QY 404 TCGCGCGGTACGCCACACCGCTTGGCTCGGCTTGGAGGATTCAGCAGGCGGAGCA 453
DB 1150 TCGCGCGGTACGCCACACCGCTTGGAGGCGAGGTCCGCGAGCTCCGCGAGCGGCCACCA 1091
QY 464 GGTGGTGGATGCGCTCGACGTCGCGGCGGAACT 500
DB 1090 GGTTCACAGCAGACCCGCTTGACCTGGGCGCCGAACTT 1054

RESULT 2
US-09-180-271-4/c
; Sequence 4, Application US/09180271
; Patent No. 6210930
; GENERAL INFORMATION:
; APPLICANT: FILIPPINI, Silvia
; APPLICANT: LOMOVSKAYA, Natalia
; APPLICANT: FONSTEIN, Leonid
; APPLICANT: COLOMBO, Anna L.
; APPLICANT: HUTCHINSON, C. Richard
; APPLICANT: OPTEN, Sharee L.
; APPLICANT: BREME, Umberto
; TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nikaido, Marmelstein, Murray & Oram
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/180,271
; FILING DATE: 14-JAN-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US98/03938
; FILING DATE: 05-MAR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/901,306
; FILING DATE: 28-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/812,412
; FILING DATE: 06-MAR-1997
; ATTORNEY/AGENT INFORMATION:
```

```
NAME: Kitts, Monica C.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P1615-8010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)638-5000
TELEFAX: (202)638-4810
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1401 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces peucetius
FEATURE:
NAME/KEY: CDS
LOCATION: 190..1398
US-09-180-271-4

Query Match      30.0%; Score 149.8; DB 3; Length 1401;
Best Local Similarity 65.3%; Pred. No. 1.2e-22;
Matches 220; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 164 GGATCCACTGTCGCCAGCGCGGCGGAACTCTCTGCTCTTCCGCAATGATCTCTCGCGT 223
DB 1390 GGATCCAGCGCGCCCTCTCTGCTGAACTCCGCTCTCGGGGCTGGACTTCTGTAGGT 1331
QY 224 GGTTCACGCGAAGAGCAGCGCTAGTCCACCGGTGCGGCGTGAACGCGTCCGGGTGC 283
DB 1330 GGTTCACGCGAAGAGCAGCGCTAGTCCGCGGTAGGCGGCCCGGAACTCTCTCGCGGAGC 1271
QY 284 GCACCGGATGTGTGTCGCGGGGTGAGCGCGCCCTCTTGGCGGGGTGCTTCGCACA 343
DB 1270 GGATCGGGATGTGGGACCCCGGGAGCCGACCGCGGATCTTGGCGGCTGTGTCTAGTA 1211
QY 344 CCCAGGAGACGAGTCCGAGCCGATGCCGAGAAAGTTCTGTCACGCTGGCGCTCTTTCGCCG 403
DB 1210 CGCAGCGCAGAGATCGGGGCCGATCCCGCATAGTTGACACTGTGGCGCTCTTGGCGG 1151
QY 404 TCGCGCGGTACGCCACACCGCTTGGCTCGGCTTGGAGGATTCAGCAGGCGGAGCA 463
DB 1150 TCGCGCGGTACGCCACACCGCTTGGAGGCGAGGTCCGCGAGCTCCGCGAGCGGCCACCA 1091
QY 464 GGTGGTGGATGCGCTCGACGTCGCGGCGGAACT 500
DB 1090 GGTTCACAGCAGACCCGCTTGACCTGGGCGCCGAACTT 1054

RESULT 3
US-09-029-603-1/c
; Sequence 1, Application US/09029603
; Patent No. 6210935
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Engel, Natalie
; APPLICANT: Bietenhader, Jurg
; APPLICANT: Toupet, Christine
; APPLICANT: Pospisch, Andreas
; TITLE OF INVENTION: Staurosporin Biosynthesis Gene Clusters
; FILE REFERENCE: 4-20555/A/PCT
; CURRENT APPLICATION NUMBER: US/09/029,603
; EARLIER FILING DATE: 1998-03-20
; EARLIER FILING DATE: 1996-08-19
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2122
; TYPE: DNA
; ORGANISM: Streptomyces longisporoflavus
; FEATURE:
; NAME/KEY: misc_RNA
```



; LOCATION: (1)..(2122)  
; OTHER INFORMATION: product = 2.1 kb region  
US-09-029-603-1

Query Match 11.4%; Score 56.8; DB 3; Length 2122;  
Best Local Similarity 66.1%; Pred. No. 0.0014;  
Matches 82; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 GGTACCGGACCGTGTCCCGGACACAGGAGTCCAGATACGGCAGAGAGAACACCCCGGT 60  
DB 271 GGGTGGCGTGGCGCCTCGGTGACGTCGTCTCTGTGTACGGCGAGACGAACTTGCCTCCCTGT 212  
QY 61 AGTCCGGGTAGACGGTGGCGGCGGAGCGTACGGCGCTTCGACCGTTCAGCGGGCGGGCGG 120  
DB 211 CGTGGGGGAGACGCGGGGGGTGAACTCGACGGCTCCCTCGACGCTGAGCGGGCGTGT 152  
QY 121 ACAC 124  
DB 151 TCAC 148

## RESULT 4

US-09-194-905-3/c  
; Sequence 3, Application US/09194905  
; Patent No. 6306627  
; GENERAL INFORMATION:  
; APPLICANT: DECKER, Heinrich  
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS  
; TITLE OF INVENTION: GLA.O AND THEIR USE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/194,905  
; FILING DATE: 29-JUL-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/EP97/02826  
; FILING DATE: 30-MAY-1997  
; PRIOR APPLICATION DATA: DE 19622783.6  
; APPLICATION NUMBER:  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granados, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 026083/0193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 546 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-194-905-3

Query Match 10.1%; Score 50.6; DB 4; Length 546;  
Best Local Similarity 49.8%; Pred. No. 0.027;  
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 239 GCAGCGGTAGTCCACCGCGTCCGGGCTGACCCGTCGGGGTCCGACCGGATGTGCG 298  
DB 383 GAGGGAGCGGTATACCTCGTCCGTGGAGACGTGTCACGAGGTCGCGACACCGTGGCGGA 324  
QY 299 TCCCGGGGTGAGCCCGCCCTGCTTGGCCCGGCGCTCGTGTGTCGACACACCCAGAGACACGAGT 358  
DB 323 GCGCGGCTCGAGCAGGACCTGGGTGCCACAGCAGCTTGGTGGCGGTGAAGCGCACCGCTGT 264  
QY 359 CCGACCGATGCGCGCAGAGTTCGTACCGTGGGGCTCTTCGGCGTCCGCGCTACGCCA 418  
DB 263 CGGTGATGGAGCGGTGACGTCGAGTTCGGCCCGGAAAGTGCACGATGTCGTGTCGCGCG 204  
QY 419 CCACCCGCTTGGCCCTCGGCCCTTGAAGCAGTTCAGCAGGGCGAGCAGTCCGTCGCGATGC 478  
DB 203 CGGCACCGTTCGACGAGCGCGTTCGCACACGTCGCCCTGGACGAAAGTGGGCGG 144  
QY 479 CCTCGACGTCCGGCGCG 495  
DB 143 GATGTCACGACCGCG 127

## RESULT 5

US-09-124-541-3/c  
; Sequence 3, Application US/09124541A  
; Patent No. 6229066  
; GENERAL INFORMATION:  
; APPLICANT: Morris Ph.D., Roy O.  
; TITLE OF INVENTION: A CYTOKININ OXIDASE  
; FILE REFERENCE: UMO1490  
; CURRENT APPLICATION NUMBER: US/09/124,541A  
; CURRENT FILING DATE: 1998-07-29  
; EARLIER APPLICATION NUMBER: 60/054,268  
; EARLIER FILING DATE: 1997-07-30  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 1605  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1605)  
US-09-124-541-3

Query Match 10.0%; Score 50.2; DB 3; Length 1605;  
Best Local Similarity 51.0%; Pred. No. 0.033;  
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

QY 165 GATCCACTGTCCGCCAGCCCGCGGAACTCTCTCTCTCGCCATGATCTCGTCCGGGTG 224  
DB 1194 GAGCCACGGGTGCGGCACCCGCCACAGCCCGCTTGTGAGCGCCACCTCTCGCGGTG 1135  
QY 225 GTTCCAGCGGAGAGCAGCGGTAGTCCACCGGTGCGGGGTGACCGTCCGGGGTGGCG 284  
DB 1134 CACCCGGTCAAGGAACCGCGTAGGCCA---CGTCGCTGGAACCGCAAC---CCCTC 1081  
QY 285 CACCGGATGTGCGTCCCGGGGTGAGCCGCGCCCTGCTTGGCCGGCTCGTGTGCGCACAC 344  
DB 1080 CAGTAGCTCAGCGTGCACGAGCAGCGAGCTCTGTGTCACCGCCCGCGCGCGC 1021  
QY 345 CCAGGAGACAGGTCCGGAACGATGCGCGAGAAATTGTCACGGTGGGCTCTTCGCGGT 404  
DB 1020 CGTGGCGTTGTCGTAGTTGAGCGTGGCTCGATCTGTACACGGTGGTGGCGTTCCGCTC 961  
QY 405 CGCGCGTACGCCACCGCCGCTTGGCCCTTGAAGGTTGAGCGAGTTCAGAGGGCGAGCAG 464  
DB 960 CCCGGCAGCGCGACGATCCGGGCGAGTGGCGGTGCGGTGAGAACCCCGTTCGCGCAG 901  
QY 465 GTCGTTGCGATGCCCTTCGAGCTCGGGCGGAACC 499  
DB 900 GTCGTCGCCAGGCTCTGGTTTCAGGAACACCGACC 866



```

; LOCATION: (3311)..(3607)
; NAME/KEY: unsure
; LOCATION: (5697)
US-09-6663-326-2

```

Query Match 10.0%; Score 50.2; DB 4; Length 6733;  
Best Local Similarity 51.0%; Pred. No. 0.033;  
Matches 171; Conservative 0; Mismatches 158; Indels 6; Gaps 2;

RESULT 9  
US-09-194-905-7/c  
; Sequence 7, Application US/09194905  
; Patent No. 6306627  
; GENERAL INFORMATION:  
; APPLICANT: DECKER, Heinrich  
; TITLE OF INVENTION: ISOLATION OF THE BIOSYNTHESIS GENES FOR  
; TITLE OF INVENTION: PSEUDO-OLIGOSACCHARIDES FROM STREPTOMYCES GLAUCESCENS  
; TITLE OF INVENTION: GLA.C AND THEIR USE  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY & LARDNER  
; STREET: 3000 K Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/194,905  
; FILING DATE: 29-JUL-1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/EP97/02826  
; FILING DATE: 30-MAY-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE 19622783.6  
; FILING DATE: 07-JUN-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granados, Patricia D.  
; REGISTRATION NUMBER: 33,683  
; REFERENCE/DOCKET NUMBER: 026083/0193  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399

```

; INFORMATION FOR SEQ ID NO: 7:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 6854 base pairs
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
US-09-194-905-7

```

Query Match 9.8%; Score 49; DB 4; Length 6854;  
Best Local Similarity 49.4%; Pred. No. 0.059;  
Matches 127; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

```

RESULT 10
US-09-252-991A-5764/C
; Sequence 5764, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELAT
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5764
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5764

```

	Query Match	9.8%;	Score 48.8;	DB 4;	Length 900;
	Best Local Similarity	47.0%;	Pred. No. 0.063;		
	Matches 185;	Conservative 0;	Mismatch 207;	Indels 2;	Gaps 1;
QY	105	GTCAAGCGGGCGGACACACGGCGGGTCAGCTCGTCAAGTACGCGGGGACGTACAG	164		
DB	528	GACGGCGCGTCGTCTCATCTGTCGGGTGCTGTTCACGAAGCAATCGCGTTCCTC	469		
QY	165	GATCCACTCTCGCGCCAGCCGGCGGAATCTCTGCTCTTCGGCCATGATCTCGTCGGCGTG	224		
DB	468	AGGCGGCAACGCTGGCGGTTCGCACCATGGGCCACTTCGC	411		
QY	225	GTTCACAGGCGAAGACAGCGCGTAGTCCACCGCGTGGCGGTGAACGGTCCGGGGTGCG	284		
DB	410	ATGCACCTCGGGAGCGCGGTGACGAATCTGTCGACGAACACCCGCAATCTCGCGCGTGAC	351		
QY	285	CACCGGGAATGTCGTGCCGGGGGTGAGCCGGCGCCCTGCTTGGCGGGCGTGGTGTGCGCACAC	344		

Db 350 TTCCTGTTCCGGTGTACCGCGGTTGTACCGCAACACCGCGGGATCGCAGCAAC 291  
QY 345 CCAGGAGACGAGTCCGACCGATGCGCAGAGTTGTACGTTGGCTCTTCGCGGT 404  
Db 290 GCCATGCGCTGCGCAGCAGGCGGCGCCATCCCGGGCGATGATGATTGTGAC 231  
QY 405 CGCAGGATACCCACACCGCTTGCCTTCGGCTTCAGGAGTTACAGAGGCGAGCAG 464  
Db 230 GTCGCGGCGCATGCGGTTGCGAGGAAACACCAACCATGATCCCCAGCAGCGTGGCGAG 171  
QY 465 GTCGAGTGGATGCTTCGAGTGGCGGCGAAC 498  
Db 170 GCCGATGAGGGCGCGCGGCGGATGGCGTTGGAC 137

RESULT 11  
US-09-252-991A-5708  
; Sequence 5708, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 5708  
; LENGTH: 2085  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5708

Query Match 9.8%; Score 48.8; DB 4; Length 2085;  
Best Local Similarity 47.0%; Pred. No. 0.064;  
Matches 185; Conservative 0; Mismatches 207; Indels 2; Gaps 1;

QY 105 GTCAGCGGGCGGCGGACACCGGGGTGACGTGTCACGTGTACGCGGGGAGCTACAG 164  
Db 189 GACGCGCGCGTGTGTCATGTCGTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTC 248  
QY 165 GATCACTGTGCGGACGCGCGGCGGAACTCTGCTCTTCCGTCATGATCTGTCGCGGTG 224  
Db 249 AGCGCGCAACGCTGCGCGGTGCGACATGGGCGCACTTCGC--TGAGCTGGCTTCGCC 306  
QY 225 GTTCCAGCGAAGACGACGCGGTAGTCCACCGCGTCGCGGCGTGAAACCGCTCCGCGGTGCG 284  
Db 307 ATGCACCTCGGCGAGGCGGTGACGAACTGTCGACGAAACACCGCGCATGTCGCGCTGAC 366  
QY 285 CACCGGATGTGCGTCCGCGGGGTGACCGGCGCTGCTTGGCGCGCGTGTGTCGACAC 344  
Db 367 TTCCTTGTTCGCGTGACGACGCGGTGTAGCCGAAACAAACCGGGGATCGCAGCAACAG 426  
QY 345 CCAGAGACGAGTCCGACCGATGCGCGGCGGCGCCATCCCGGGCGGATGCGATTGATTGAC 404  
Db 427 GCCATGCGCTGCGCAGCAGGCGGCGCCATCCCGGGCGGATGCGATTGATTGAC 486  
QY 405 CGCGCGGTACGCGACCGCCCTTGCCTTCGGCTTTAGCGAGTTACGAGGCGGAGCAG 464  
Db 487 GTCGCGGCGCATGCGGTGCGAGGAAACACCAACCATGATCCCCAGCAGCGTGGCGAGCAG 546  
QY 465 GTCGAGTGGATGCGCTCGAGTGGCGGCGAAC 498  
Db 547 GCCGATGAGGGCGCGCGGCGGATGGCGTTGGAC 580

## RESULT 12

US-08-852-401-1/c

; Sequence 1, Application US/08852401  
; Patent No. 5976836  
; GENERAL INFORMATION:  
; APPLICANT: Weber, J. Mark  
; APPLICANT: Hessler, Paul E.  
; APPLICANT: Larsen, Peter E.  
; APPLICANT: Luu, B. Minh  
; TITLE OF INVENTION: Methods and Compositions for Enhancing  
; TITLE OF INVENTION: Erythromycin Production  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rockey, Milnamow & Katz, Ltd.  
; STREET: 2 Prudential Plaza, 180 N. Stetson, Suite  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,401  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueller, Lisa L.  
; REGISTRATION NUMBER: 38,978  
; REFERENCE/DOCKET NUMBER: FER2159POO300US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-616-5400  
; TELEFAX: 312-616-5460  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4776 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 9.8%; Score 48.8; DB 2; Length 4776;  
Best Local Similarity 47.4%; Pred. No. 0.064;  
Matches 146; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 192 CTCCTGCTCTTCGCGCATGATCTCGTCGCGGTGTTCCAGGGAAGACAGCGGTAGTC 251  
Db 1644 CACCGCGCGCGCGCATCTCTCGATCGCTGCGCCACGCTCGTCAGCGCGGTGGCA 1585  
QY 252 CACCGCGTCGCGGTGAACCGGTGCGCACCGGATGTCGTCGCGGGGTGAG 311  
Db 1584 GTAGAGGCCAGCGCATGTCGTGAAGCCGACACCGACAGGTCTCTCGGGGTGGCCAG 1525  
QY 312 CCGGCGCTGCTTGGCGCGGTGTCGTGTCGACACCGCAGAGACCGAGTCCGACCGATGCC 371  
Db 1524 TCCGCGCGCTTGGCTCGCATCGCGCATCGCGCGCCATGACGTCCGAGCAGCATCAC 1465  
QY 372 GCAGAGTTCTGTCAGGTGGCGCTCTTCGCGCTGCGCGCTACGCCACACCGCGTTGCC 431  
Db 1464 GCGGTGGGCGCGGTTCGAGCAGCTCGGCAATGSCCTTCGCGCGCGCTTCGCGCGCC 1405  
QY 432 CTCGCGCTTGAAGAGTTTCAGCAGGCGGAGCAGGTTCGTTGCGGATGTCCTCGAGTCCGC 491  
Db 1404 GAACGCGCGGTGCGCGACGACCGCGGTTCGCGCGGATGTCGCGCTCTCTCCAGCGCAGC 1345  
QY 492 GCGGAC 499  
Db 1344 GCGCCAGC 1337

## RESULT 13

```

US-08-911-853-3/c
; Sequence 3, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Garrise, Gijbsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,099
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1416 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-911-853-3

```

Query Match	9.7%;	Score 48.6;	DB 3;	Length 1416;
Best Local Similarity	49.8%;	Pred. No. 0.069;		
Matches 154;	Conservative	0;	Mismatches 149;	Indels 6; Gaps 1;
QY	188	GGAACTCCTGCTCCTTCGCTCCCATGATCTCGTCGGGGTGTTCCAGGCGGAAGACGACGGCGT	247	
Db	319	GGAGCATCTCGTCGTGATCGAAGGECTTGGCGATGTAGTCCAGCGCGCCATCTTTCATCG	260	
QY	248	AGTCACACCGCGTCGGGCGGTGAACGCGTCCGGGTGCGCACCGGGATGTCGTCGCCGGGG	307	
Db	259	AGTCACACCGCGCGAGCGAGGTGCGCATAGTGTGTCATGATCAGTACCGGGGTGCCGTGCG	200	
QY	308	TGAGCGGGCCCTGTCTTGGTCGGCGCTCGGTGTGCACACCCAGGAGACGAGTCCGACCGGA	367	
Db	199	CCAGCTTGATCAGCTCGGTGCGGGGGCGCGGGCA-----GGCGCAGTTCGCTGACCA	146	
QY	368	TGCCGACAGATTCTGTACGGTGGCGTCTTTCGCCGTCGGCGCGTAGGCCACACCCCGCT	427	
Db	145	CCAGTCCGAAGGTCCGAATGTGTGTAGCGCTCCTTGGGCTCTCTTGAACCGAACCGGCTCGC	86	
QY	428	TGCCCTTCGGCCCTTGAGCGAGTTTCAGCAGGGCGAGCAGGTGGTGGCGGATGCCCTCGAGCT	487	
Db	85	TGACTGTGTACTGGTTGCGTTCAGACAGGGCGGCGCAGGGCGGAGCGGATGATGGTTTCGT	26	
QY	488	CGGCGGCGGA	496	
Db	25	CTTCGACGA	17	

```

/ Sequence 3, Application US/09479409
/
/ Patent No. 6225106
/ GENERAL INFORMATION:
/
/ APPLICANT: Gerritts, Gijabert
/ APPLICANT: Quax, Wilhelms J.
/
/ TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
/ TITLE OF INVENTION: EXPRESSION LEVELS
/
/ NUMBER OF SEQUENCES: 37
/
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Genencor International
/ STREET: 925 Page Mill Road
/ CITY: Palo Alto
/ STATE: CA
/
/ COUNTRY: USA
/ ZIP: 94304-1013
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/
/ SOFTWARE: Fast-Seq for Windows Version 2.0
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/479,409
/
/ FILING DATE:
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/911,853
/
/ FILING DATE:
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Glaister, Debra J
/ REGISTRATION NUMBER: 33,888
/ REFERENCE/DOCKET NUMBER: GC361-2
/
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 650-846-7620
/ TELEFAX: 650-845-6504
/
/ INFORMATION FOR SEQ ID NO: 3:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1416 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ PS-09-479-409-3

```

Query Match	9.7%	Score 48.6	DB 3	Length 1416
Best Local Similarity	49.8%	Pred. No. 0.069		
Matches 154	Conservative 0	Mismatches 149	Indels 6	Gaps 1
Qy	188	GGAACTCTGCTCCTTCGCCATGATCTCGTCGGCGTGGTTCAGGCGAAGACGACGCGGT	247	
Db	319	GGAGCATCTCGTCGTGATCGAAGGGCTGGCGATGAGTCCACGCGCCCATCTTCATCG	260	
Qy	248	AGTCCACCGCGTCCGGGCGTGAACGCGTCCGGGGTGCACCGGGATGTGGTTCGCGGGG	307	
Db	259	AGTCCACCGCGAGCGAGCTGGCATAGTGGTCATGATCAGTACCGGGGTGCGTCGG	200	
Qy	308	TGAGCGCGGCCCTGCTTGGCGCGCGTGTGTGTCACACCCAGBAGACAGAGTTCGAGACGA	367	
Db	199	CCAGCTGTGATCAGCTCGGTGTCGGGGGCGCGGGCA-----GGCGCAGGTTCGCTGACCA	146	
Qy	368	TGCCGCAGAGTTTCGTCACGTTGGCGCTCTTCGCGTCGCGCGCTACGCCACACCCGCT	427	
Db	145	CCAGGTCGAAGGTCGGAATCTGTAGCGCTCCTTGCGCCTCCTGAACCGAACCGCGCTGCG	86	
Qy	428	TGCGCTTCGGGCTTTGAGCGAGTTTCAGCAGGGCGAGCAGGTTCGGTCGGATGCGCTTCGACGT	487	
Db	85	TGACCTGGTACTGGTTGCGGTTCCAGCAGCGCGCGCAGGGCGAGCGGATGATGGTTTCGT	26	
Qy	488	CGCGCGCGCA	496	
Db	25	CTTCGACGA	17	

Patent No. 6313283  
GENERAL INFORMATION:  
APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,453  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J.  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1416 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-453-3

Query Match 9.7%; Score 48.6; DB 4; Length 1416;  
Best Local Similarity 49.8%; Pred. No. 0.069;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAATCTCTGCTTCCATGATCTCGTCCGCGTGGTCCAGCGAAGACGCGGT 247  
Db 319 GGAGCATCTCGTGTGATCGAAGGCTTGGCGATGTAGTCCACCGCGCCATCTTCATCG 260  
QY 248 AGTCCACCGCTCGGGGTGACCGGTCCGGGTCCGACCGGATGTGGTCCGGGG 307  
Db 259 AGTCCACCGCGAGCGCGAGGTGGCATAGTGGTGTATGATCAGTACCGGGTCCGCG 200  
QY 308 TGAGCCGGCCCTGCTTGGCCGCGTCTGTGCGCACCCAGGAGACGAGTCCGACCGA 367  
Db 199 CCAGCTTGATCAGTCTGCTGCGGGGCGCGGCA-----GGCGAGGTCTGCGTACCA 146  
QY 368 TGCCGACAGTCTGTCAGGTGGCGCTTTCGCGGTCCGCGGTGCGCACACCGCT 427  
Db 145 CCAGTCAAGGTGCGAATGTGTAGCGCTCTGGGCTCTCTGAACCGAACCGGCTCGC 86  
QY 428 TGCCCTCGGCTTGAGCGAGTTTCAGAGGGGAGCAGGTCCGTCGCGATGCCCTCGAGT 487  
Db 85 TGACCTGTACTGTTGGTTCCTCAGCAGCGCGCGAGGCGAGCGGATGATGTTTCGT 26  
QY 488 CGCGCGCGA 496  
Db 25 CTTGACGA 17

RESULT 16  
US-09-911-853-28/c  
Sequence 28, Application US/08911853  
Patent No. 6048710

GENERAL INFORMATION:  
APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911,853  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/699,092  
FILING DATE: 16-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J.  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-911-853-28

Query Match 9.7%; Score 48.6; DB 3; Length 4377;  
Best Local Similarity 49.8%; Pred. No. 0.071;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAATCTCTGCTTCCATGATCTCGTCCGCGTGGTTCAGCGGAGACGCGGT 247  
Db 1761 GGAGCATCTCTGCTGTGATCGAAGGCTTGGCGATGTAGTCCACCGCGCCATCTTCATCG 1702  
QY 248 AGTCCACCGCTCGGGGTGAAACCGGTCCGGGTGCGCACCGGATGTGCGTCCGGGG 307  
Db 1701 AGTCCACCGCGAGCGAGCTGGCATAGTGGTATGATCAGTACCGGGTCCGTCGG 1642  
QY 308 TGAGCCGGCCCTGCTTGGCCGCGTCTGTGCGCACCCAGGAGACGAGTCCGACCGA 367  
Db 1641 CCAGCTTGATCAGTCTGCTGCGGGGCGCGCGCA-----GGCGAGGTCTGCTGACCA 1588  
QY 368 TGCCGACAGTTCGTCACGCTGCGCTCTTCCGCGTCCGCGGTACGCCACACCGCT 427  
Db 1587 CCAGTTCAGGTGCGAATGTGTAGCGCTCTTGGGCTCTCTGAACCGAACCGGCTCGC 1528  
QY 428 TGCCCTCGGCTTGAGCGAGTTTCAGCGGCGAGCGGTCCGTCGCGATGCCCTCGAGT 487  
Db 1527 TGACCTGTACTGTTGGTTCCTCAGCAGCGCGCGAGCGGATGATGTTTCGT 1468  
QY 488 CGCGCGCGA 496  
Db 1467 CTTGACGA 1459

RESULT 17  
US-09-479-409-28/c  
Sequence 28, Application US/09479409  
Patent No. 6225106  
GENERAL INFORMATION:

APPLICANT: Gerritse, Gijbert  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,409  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-409-28

Query Match 9.7%; Score 48.6; DB 3; Length 4377;  
Best Local Similarity 49.8%; Pred. No. 0.071;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAACTCTGCTCTTGGCCATGATCTCGTGGCGTGGTTCAGCGAAGACGCGGT 247  
DB 1761 GGAGCATCTCGTGTGATCGAAGGCTTGGCGATGTAGTCCACCGGCCCATCTTCATCG 1702  
QY 248 AGTCCACCGCTCGGCGTGAACGCGTCCGGGTGGCACCGGATGCGTGCCTGGCGG 307  
DB 1701 AGTCCACCGCGAGCGAGCTGGCATAGCTGTGATGATAGTACCGGGTGGCTCGG 1642  
QY 308 TGAGCGCGCTCTGTTGGCGGCGTGTGTGTCACACCCAGGAGACCGGTCCGACCGA 367  
DB 1641 CCAGCTTGATCAGCTCGTGGCGGGGCGCGGCA-----GGCGAGGTGCTGACCA 1588  
QY 368 TGCGCGAGAGTTCGTACGCTGCGGTGCGCTCTTCCGCTGCGCGCTACGCCACCCCGCT 427  
DB 1587 CCAGGTGGAAGTTCGGAATGCTGTAGCGCTCTTGGGCTCTTGAACCGAACCGGCTCGC 1528  
QY 428 TGCCCTCGGCTTTGAGCGAGTTTCAGCAGGCGGAGCAGGTGCGTGCAGATCCCTCGACGT 487  
DB 1527 TGACCTGGTACTGTTGCGTTCCAGCAGCGGCGGCGAGCGGATGATGTTTCGT 1468  
QY 488 CGGCGCGCA 496  
DB 1467 CTTGACGA 1459

RESULT 18  
US-09-479-453-28/C  
; Sequence 28, Application US/09479453  
; Patent No. 6313283  
; GENERAL INFORMATION:  
; APPLICANT: Gerritse, Gijbert

APPLICANT: Quax, Wilhelmus J.  
APPLICANT: Quax, Wilhelmus J.  
TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED  
TITLE OF INVENTION: EXPRESSION LEVELS  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genencor International  
STREET: 925 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1013  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/479,453  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/911,853  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Glaister, Debra J  
REGISTRATION NUMBER: 33,888  
REFERENCE/DOCKET NUMBER: GC361-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-846-7620  
TELEFAX: 650-845-6504  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-479-453-28

Query Match 9.7%; Score 48.6; DB 4; Length 4377;  
Best Local Similarity 49.8%; Pred. No. 0.071;  
Matches 154; Conservative 0; Mismatches 149; Indels 6; Gaps 1;  
QY 188 GGAACTCTGCTCTTGGCCATGATCTCGTGGCGTGGTTCAGCGAAGACGCGGT 247  
DB 1761 GGAGCATCTCGTGTGATCGAAGGCTTGGCGATGTAGTCCACCGGCCCATCTTCATCG 1702  
QY 248 AGTCCACCGCTCGGCGTGAACGCGTCCGGGTGGCACCGGATGCGTGCCTGGCGG 307  
DB 1701 AGTCCACCGCGAGCGAGCTGGCATAGCTGTGATGATAGTACCGGGTGGCTCGG 1642  
QY 308 TGAGCGCGCTCTGTTGGCGGCGTGTGTGTCACACCCAGGAGACCGGTCCGACCGCA 367  
DB 1641 CCAGCTTGATCAGCTCGTGGCGGGGCGCGGCA-----GGCGAGGTGCTGACCA 1588  
QY 368 TGCGCGAGAGTTCGTACGCTGCGGTGCGCTCTTCCGCTGCGCGCTACGCCACCCCGCT 427  
DB 1587 CCAGGTGGAAGTTCGGAATGCTGTAGCGCTCTTGGGCTCTTGAACCGAACCGGCTCGC 1528  
QY 428 TGCCCTCGGCTTTGAGCGAGTTTCAGCAGGCGGAGCAGGTGCGTGCAGATCCCTCGACGT 487  
DB 1527 TGACCTGGTACTGTTGCGTTCCAGCAGCGGCGGCGAGCGGATGATGTTTCGT 1468  
QY 488 CGGCGCGCA 496  
DB 1467 CTTGACGA 1459

RESULT 19  
US-09-818-780-67/C  
; Sequence 67, Application US/09818780  
; Patent No. 6671146  
; GENERAL INFORMATION:  
; APPLICANT: McHenry, Charles  
; TITLE OF INVENTION: NOVEL THERMOPHILIC POLYMERASE III HOLOENZYME

```
; FILE REFERENCE: 1794_0030004
; CURRENT APPLICATION NUMBER: US/09/818,780
; CURRENT FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/192,736
; PRIOR FILING DATE: 2000-03-28
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 2082
; TYPE: DNA
; ORGANISM: Thermus thermophilus
; US-09-818-780-67

Query Match          9.6%; Score 48.2; DB 4; Length 2082;
Best Local Similarity 48.1%; Pred. No. 0.084;
Matches 137; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 182 CCGCGGGAAGTCTGCTCTTCCGCGATGATCTGCGGCGTGTTCAGGCGAAGACA 241
Db 1558 CCAGGAAGTCTCGAGGTCTCCGCTCTTGGCGGCTCTGAGGAGCTCTCCACGTTCT 1499
QY 242 GCGGTAAGTCCACCGCGTGGGGGTGAACGCGTCCGGGGTGGCGACCGGGATGTGGTGC 301
Db 1498 CCAGGCGGTCTCGGCGTCTCGGGGTAGGCTCCCGGAGGTAGGCGGGTAGTCTGGTGG 1439
QY 302 CGGGGGTGAGCGGCGCTCTGTTGGCGCGGCTGTGTGCGCACACCCAGGAGACAGTCCG 361
Db 1438 CCTCGAGGAGTGGCGGAAGAGGCTTCGCGGGCGCGAAGACAGGTCTCTGAAGTCTCT 1379
QY 362 GACCGATGTCGCGCAGAGTTCGTACGGTGGCGCTCTTCGCGCGTGGCGCGTACGCCACCA 421
Db 1378 CCACAGGCGCACGAAATGTCTCAGGGGTCTGGCGGGGGAAGTCTTGGCGGCTCTCT 1319
QY 422 CCCGCTTGGCGTCCGCTTGGCGAGTTCAGCGAGTTCAGGAGGCGAGCAGGT 466
Db 1318 TCAGGGCTCTCCAGGAGGAAGCCCTTCTCCTGGGCGAGAGCT 1274

RESULT 20
US-08-074-121-4/c
; Sequence 4, Application US/08074121
; Patent No. 5767362
; GENERAL INFORMATION:
; APPLICANT: Best, Elaine
; APPLICANT: Knauf, Vic C.
; TITLE OF INVENTION: Methods and Compositions for Modulating
; TITLE OF INVENTION: Lipid Content of Plant Tissues
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/074,121
; FILING DATE: 08-JUN-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Rae-Venter, Barbara
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: 05938/043001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-5277
; TELEFAX: (415) 854-0875
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3231 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 861..1328
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1349..2695
; US-08-074-121-4

Query Match          9.6%; Score 48; DB 1; Length 3231;
Best Local Similarity 48.8%; Pred. No. 0.093;
Matches 159; Conservative 0; Mismatches 165; Indels 2; Gaps 1;

QY 47 GAACACCCCCCGGTAGTCCGGGTAGACGGTGGGCGGAGCGTACGCGCTTCGACGGT 106
Db 1822 GATGATACACCGGGTAGCGACCTCGCGGGCGATCGCAGGGCGGTCTCTTCATCTTCGG 1763
QY 107 CAGCGGGGGCGCGGACACCGG--CGGGTCAGCTCGTCACTGTACGCGGGGAGCGTACAG 164
Db 1762 CAGCGGGCGGTGCGAGCCCGGACCGTGGGGAACGCGCGCGCTTCATGGGTCCTTGGC 1703
QY 165 GATCACAATGTCCGCCAGCCCGCGGAACTCTGTCTCTTCCCATGATCTCTGCGGGTG 224
Db 1702 CGAAACCTTGTCGCCCATCAGGCGGATCACCTCGCGCGGTCCGGGCGGAGGTAAGCC 1643
QY 225 GTTCCAGCGGAGAGCAGCGGTAGTCCACCGCGTGGGGCGTGAACCGGTCCGGGGTGC 284
Db 1642 GGAGGTTTCATCTGCTCGCGGAAGTCCGCGCTTCTCGCGAGGAGCGGTAGCGGGTG 1583
QY 285 CACCGGAGTGTGTCGCGCGGGGTAGCGGCGCTCTTGGCGCGGCTGTGCTGTCGACAC 344
Db 1582 GATCGCGGTGGCGCGGTGACCTCGCGCGCGCGATGATGATCGCGGATCTCGAGTACGA 1523
QY 345 CCAGGAGACCGAGTCCGACCGATGC 370
Db 1522 CTGGGTGGCGGGCGGACCGATGC 1497

RESULT 21
PCT-US94-06447-4/c
; Sequence 4, Application PC/TUS9406447
; GENERAL INFORMATION:
; APPLICANT: Calgene, Inc.
; TITLE OF INVENTION: Methods and Compositions for Modulating
; TITLE OF INVENTION: Lipid Content of Plant Tissues
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weil, Gotshal & Manges
; STREET: 2882 Sand Hill Road, Suite 280
; CITY: Menlo Park
; STATE: California
; COUNTRY: US
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/06447
; FILING DATE: 06-JUN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rae-Venter, Barbara
; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: CGNE-097/WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 926-6200
```



TELEFAX: (415) 854-3713  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3231 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 861..1328  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1349..2695  
PCT-US94-06447-4

Query Match 9.6%; Score 48; DB 5; Length 3231;  
Best Local Similarity 48.8%; Pred. No. 0.093;  
Matches 159; Conservative 0; Mismatches 165; Indels 2; Gaps 1;

QY	47	GAACACCCCGGCTAGTCCGGGTAGACGGGTGGCGCGGAAAGGGGTACGCGCCCTTCGACGGT	106
DB	1822	GATGATCACCGGGTAGCGACCTCGCGGGGATCGCCAGGGGGTCTCTTCATCTTCGG	1763
QY	107	CAGCGGGGGCGGACACCGG--CGGGTACGCTCGTCACGTGTACGCGGGGACGTACAG	164
DB	1762	CAGCGGGCGGTGCGAGCCCGGACGCGTGGGGACGCGCGCGCGCTTCATGCGCTCCTTGGC	1703
QY	165	GATCCACTGTCCGCCAGCCCGGAACTCCTCTCTCGCCCATGATCTCGTGGCGGTG	224
DB	1702	CGAACTTGTCCGCCCATCAGCGGGATCACCTCGGGGTCGGGCCGACGAGGTGAAGCC	1643
QY	225	GTTCCAGGGAAGAGACGCGGTAGTCCACCGCGTGGGGGTGAACGCGTCCGGGGTGG	284
DB	1642	GGAGCGTTCGATCTGTCGCGAAGTCGGCGTTCCTCGGGGAGGAAGCCGTAGCGGGGTG	1583
QY	285	CACCGGATGTGGTGGCGGGGTGAGCGCGCCCTCTTGGCGGGGTGCTGTGCGCACAC	344
DB	1582	GATCGGGTGGCGCGGTGACCTCGCGCGGGGATGATCGCGGGATTCGAGGTACGA	1523
QY	345	CCAGGAGACAGTCCGACCGATGC	370
DB	1522	CTGGGTGGCGGGCGCGACCGATGC	1497

RESULT 22  
US-09-635-132-3/c  
; Sequence 3, Application US/09635132  
; Patent No. 6620601  
; GENERAL INFORMATION:  
; APPLICANT: YAMAGUCHI, ISAMU  
; APPLICANT: NAKASHITA, HIDEO  
; APPLICANT: YOSHIOKA, KEIKO  
; APPLICANT: DOI, YOSHIHARU  
; TITLE OF INVENTION: METHODS FOR TRANSFORMATION OF PLANTS, TRANSFORMED  
; FILE REFERENCE: 081356/0148  
; CURRENT APPLICATION NUMBER: US/09/635,132  
; PRIOR FILING DATE: 2000-08-09  
; PRIOR APPLICATION NUMBER: JP 11-225832  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: JP 11-225839  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1181  
; TYPE: DNA  
; ORGANISM: Ralstonia eutropha  
US-09-635-132-3

Query Match 9.6%; Score 47.8; DB 4; Length 1181;

RESULT 23  
US-08-241-943-23/c  
; Sequence 23, Application US/08241943  
; Patent No. 5602321  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E.  
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS  
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS BIOPOLYESTIC  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
; STREET: First Wisconsin Plaza, One South  
; STREET: Pinckney St.,  
; STREET: P.O. Box 2113  
; CITY: Madison  
; STATE: WI  
; COUNTRY: USA  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/241,943  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION NUMBER: us/07/980,521  
; FILING DATE: 20-NOV-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27,386  
; REFERENCE/DOCKET NUMBER: 11-229-9076-8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (608) 251-2484  
; TELEFAX: (608) 251-9166  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1182 base pairs  
; TYPE: nucleic acid

```
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; HYPOTHETICAL: NO
; PUBLICATION INFORMATION:
; AUTHORS: PEOPLES
; JOURNAL: J. Biol. Chem.
; VOLUME: 264
; PAGES: 15293-15297
; DATE: 1989
;
US-08-241-943-23
Query Match
Best Local Similarity 9.6%; Score 47.8; DB 1; Length 1182;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCAGCGGGCGGCGGACACCGGGGTCAGCTGTCAGCTGTCAGCGGGGACGTACAG 164
DB 411 GGCATCGGCCATGCGGAACCATCGCGAGCCCGGACAGACGTGGGGGGCGGCTCAT 352
QY 165 GATCCACTGTCCGCGAGCCCGGGAACCTCTGCTTCCTTCGCCATGATCTGTCGGGT- 223
DB 351 GTTTTCTGGCGCGGCCACCAACAGATCTCGCGGTGCGGCCGCCATGATCGGTCG 292
QY 224 --GGTTCAGGGAAGACGAGCGGCTAGTCAACCGCTCGGGCTGACGCGGTGGGGT 281
DB 291 CAGCATCAGCGCTTCAGCGCCGAGCCGACACCTTGTGATGTCATGCGCGCACCAT 232
QY 282 GCGCACCGGATGTCGTGCGGGGGTGAGCGGCGCTTGTGCGCGGCTGTCGTGCGCA 341
DB 231 CCGCGGAGCGCGCTTGATCGCGCTGCGGTGCGGGTTCGCGCCGAAACCGGGT 172
QY 342 CACCCAGAGACCGAGTCGGGACCGATCGCGGAGGAGTTCGTCAGGTGGGCTTCGCG 401
DB 171 CAGCACCTGCGCCATGACTTCGCTCACTGCTCGGCTTGCAGCGCGCGGCTCCAG 112
QY 402 CCGCGCGGTGAGCCACCGCCGCTGCGCTTCGCGCTTGCAGGAGTTCAGCAGGGCGAG 461
DB 111 CCGGCGCTTGATGACCAACCGGACCGGACCGGTCGCGTTCGCGGTCGCGGTCG 52
QY 462 CAGGTGCGGTGCGGATGCGCTCGACGTGCGGCGCA 496
DB 51 AAACCTGCGGACCGGCGGTGCGGCGCGGATACGA 17

RESULT 24
5229279-5/c
; Patent No. 5229279
; APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
; TITLE OF INVENTION: METHOD FOR PRODUCING NOVEL POLYESTER
; BIOPOLYMERS
;
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/556,535
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-AUG-1987
; SEQ ID NO:5:
; LENGTH: 2327
5229279-5

Query Match
Best Local Similarity 9.6%; Score 47.8; DB 6; Length 2327;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCAGCGGGCGGCGGACACCGCGGGTCACTGTCGTCAGCGGGGACGTACAG 164
DB 450 GGCATCGGCCATCGGAAACCATCGCGCGAGCCCGGACGACGTCGCGGGCGGCTCAT 391
QY 165 GATCCACTGTCCGCGAGCGCGGGAACCTCTGCTTCCTTCGCCATGATCTGTCGGCGT- 223
DB 390 GTTTTCTGGCGCGCGGCCACACAGATCTCGGCGTTCGCCGCTCATGTCGGGCGG 331
QY 224 --GGTTCAGGGAAGACGAGCGGTCAGTCAGCGGTGCGGGGTGAAACGCTTCGGGT 281
DB 330 CAGCATCAGCGCTTCAGGCGCGGACGACACCTTGTGATGTCATGCGCGCGGACCAT 271
QY 282 GCGCACCGGATGTCGTGCGGGGTGAGCGGCGCTCTGCTTCGCGCGGTGTCGTCGCA 341
DB 270 CGCGCGGAGCGCGCTTGTATCGCGCTGCGGTGCGGGTTCGCGCCGAAACCGGGT 211
QY 342 CACCCAGGAGACGAGTTCGCGACCGGATGCGCGAGAGATTCGTCAGGTGGGCTTCGCG 401
DB 210 CAGCACCTGCGCCATGATGACTTCGCTCACTGCTCGGCTTCGCGCGCGGCTCCAG 151
QY 402 GTCGCGCGGTGAGCCACCGCCGCTTGCCTTCGCGCTTGCAGGAGTTCAGCAGGGCGAG 461
DB 150 CGCGGCTTGTATGATACCAACCGGACCGGATTCGCGGTGCGGGATCTTGCGCGAGCGCGCC 91

RESULT 25
5512669-5/c
; Patent No. 5512669
; APPLICANT: PEOPLES, OLIVER P.; SINSKEY, ANTHONY J.
; TITLE OF INVENTION: GENE ENCODING BACTERIAL ACETOACETYL-COA
; REDUCTASE
;
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/297,667
; FILING DATE: 29-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 124,570
; FILING DATE: 20-SEP-1993
; APPLICATION NUMBER: 944,488
; FILING DATE: 03-NOV-1992
; APPLICATION NUMBER: 566,535
; FILING DATE: 13-AUG-1990
; APPLICATION NUMBER: 67,695
; FILING DATE: 29-JUN-1987
; SEQ ID NO:5:
; LENGTH: 2327
5512669-5

Query Match
Best Local Similarity 9.6%; Score 47.8; DB 6; Length 2327;
Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY 105 GTCAGCGGGCGGCGGACACCGCGGGTCACTGTCGTCAGCGGGGACGTACAG 164
DB 450 GGCATCGGCCATCGGAAACCATCGCGCGAGCCCGGACGACGTCGCGGGCGGCTCAT 391
QY 165 GATCCACTGTCCGCGAGCGCGGGAACCTCTGCTTCCTTCGCCATGATCTGTCGGCGT- 223
DB 390 GTTTTCTGGCGCGCGGCCACACAGATCTCGGCGTTCGCCGCTCATGTCGGGCGG 331
QY 224 --GGTTCAGGGAAGACGAGCGGTCAGTCAGCGGTGCGGGGTGAAACGCTTCGGGT 281
DB 330 CAGCATCAGCGCTTCAGGCGCGGACGACACCTTGTGATGTCATGCGCGCGGACCAT 271
QY 282 GCGCACCGGATGTCGTGCGGGGTGAGCGGCGCTCTGCTTCGCGCGGTGTCGTCGCA 341
DB 270 CGCGCGGAGCGCGCTTGTATCGCGCTGCGGTGCGGGTTCGCGCCGAAACCGGGT 211
QY 342 CACCCAGGAGACGAGTTCGCGACCGGATGCGCGAGAGATTCGTCAGGTGGGCTTCGCG 401
DB 210 CAGCACCTGCGCCATGATGACTTCGCTCACTGCTCGGCTTCGCGCGCGGCTCCAG 151
QY 402 GTCGCGCGGTGAGCCACCGCCGCTTGCCTTCGCGCTTGCAGGAGTTCAGCAGGGCGAG 461
DB 150 CGCGGCTTGTATGATACCAACCGGACCGGATTCGCGGTGCGGGATCTTGCGCGAGCGCGCC 91
```



; APPLICANT: YAWAGUCHI, ISAMU  
 ; APPLICANT: NAKASHITA, HIDEO  
 ; APPLICANT: YOSHIOKA, KEIKO  
 ; APPLICANT: DOI, YOSHIHARU  
 ; TITLE OF INVENTION: METHODS FOR TRANSFORMATION OF PLANTS, TRANSFORMED  
 ; TITLE OF INVENTION: PLANTS AND PROCESSES FOR PREPARATION OF POLYESTERS  
 ; FILE REFERENCE: 081356/0148  
 ; CURRENT APPLICATION NUMBER: US/09/635,132  
 ; CURRENT FILING DATE: 2000-08-09  
 ; PRIOR APPLICATION NUMBER: JP 11-225832  
 ; PRIOR FILING DATE: 1999-08-09  
 ; PRIOR APPLICATION NUMBER: JP 11-225839  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4984  
 ; TYPE: DNA  
 ; ORGANISM: Ralstonia eutropha  
 ; US-09-635-132-1

Query Match 9.8%; Score 47.8; DB 4; Length 4984;  
 Best Local Similarity 46.8%; Pred. No. 0.1;  
 Matches 185; Conservative 0; Mismatches 207; Indels 3; Gaps 1;

QY	105	GTCAAGCGGGCGGCGGACACCGCGGTTCAGCTCAGCTGTACGGCGGGGACGTACAG	154
Db	3106	GGCATGTGCCCATTCGGGAAACATTCGCGGAGCCCGGACGACGTGCGGGCGGCGTTCAT	3047
QY	165	GATCCACTGTCCGCCAGCCCGCGGGAACCTCTGCTCTTCGCCATGATCTCGTGGCGGT	223
Db	3046	GTTTTCTCGCGCGCGGCCACCACTCTCGGCGGTTCGCCGCATGATCGTTCGCGCGC	2987
QY	224	--GTTTCCAGGAGACGACGCGTAGTCCACCGCGTCGGGCGTGACCGGTCCGGGT	281
Db	2986	CAGCATCAGGCGCTTCAGGCGCGGACGCGACACCTTGTGATGTCATGCCCGCACCAT	2927
QY	282	CGCCACCGGGATGTCCGTGCGGGGGGTGAGCGGCGCTCTGTGGCCCGCGCTCGTGTCCGA	341
Db	2926	CGCGGCGAGCCGCGCTTGATCGCGGCTTGGCGTTCGGGGTTCCTGGCCCCGAACCGCGGT	2867
QY	342	CACCAGGAGACAGGTTCGGACCGATGCGCGAGAAATTGTCACGTTGGCGCTCTTCGC	401
Db	2866	CAGCACCTGGCCCATGATGACTTCGCTCACTGCTCCGGGTTCAGCGCGCGCTCCAG	2807
QY	402	CGTCGCGCGGTACGCCACCCCGGCTTGCCTCGGCGCTTGAGCGAGTTCAGCAGGGCGAG	461
Db	2806	CGCGGCGCTTGATGACCAAGCGCACCCAGTTCGGTTCGCGGATCTTGCCACGAGCGGCC	2747
QY	462	CAGTCCGGTCCGGATGCCCTTGACGTCCGGCGGA	496
Db	2746	AACTTTCGCACCGCGGTGCGGGCGGGGATACGA	2712

```

RESULT 29
US-09-252-991A-13466/c
; Sequence 13466, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: MARC J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13466
; LENGTH: 720
; TYPE: DNA

```

```

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13466

Query Match          9.4%; Score 47; DB 4; Length 720;
Best Local Similarity 45.3%; Pred. No. 0.15;
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 114 CGGCGGACACCGCGGGTCAGCTCGTTCAGTGTACGCGGGGACGTACAGGATCCACTG 173
DB 638 CCGCAGGACACCTCGAGGACGATCGCCAGGCGCACGCCGCTGGCGAGGATCCACAGGGT 579

QY 174 TCGCCAGACCGCGGGGAATCTCTGCTTCCTGCGCATGATCTCGTCGCGGTGTTCACAGC 233
DB 578 GTCGAACCGCGCGCTTGGGCAACCAACGCGGTCTGTAGACCTGTGATCGCGAACAGTCCCGAGC 519

QY 234 GAAGAGCAGCGGTAGTTCACCGCGTCGCGCGGTGAACCGTTCGCGGGGTGCGCACCGGAT 293
DB 518 GACCGCAGACGCTTGGCCACCAACGCGCGGCGGTGACCTCGCGGTAGGAACGCCAGAG 459

QY 294 GTGCGTCCCGGGGTGAGCCGCGCCCTCTTGGCGCGCGGTCTGTGCGACACCCAGGAGAC 353
DB 458 CCGCTTGAGCGGGCCGGAAGAACCAACGATGTTCCGGCAAGCGCTGCGCTAGTGCAGCCCG 399

QY 354 CAGTCCCGGACCGATGCGCGAGAAAGTTCGTTCAGGTGCGCGCTCTTCGCCCGTCGCGCGGTA 413
DB 398 CCGCTCGCGCGGTAGCGGCACCTGGCGAAGACCGCGCTGTCAGCGCTTCCAG 339

QY 414 GCGCACACCGCTTGCCCTCGCGCTTGACGAGATTACAGAGGCGGAGCAGGTGCGTGCG 473
DB 338 CTCGGCAGCGGATCTCTGGGTTCGCGCTCGCTGCGGTACCAAGCACCTCGCGGTG 279

QY 474 GATGCCCTCGACGTC 488
DB 278 CTCGCCGTGTTGCC 264

RESULT 30
US-09-252-991A-13771
; Sequence 13771, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13771
; LENGTH: 1992
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13771

```

QY 294 GTGCTGCGGGGGTGGAGCGGCCCTTGGCCGGCTGTGTCACACACCGAGAGAC 353  
Db 1411 CGCCTTGGCGGCGGAGACAGTGTTCGGGCGAGCGCTGGGTGTGCGCCACCG 1470  
QY 354 CAGGTCGAGCAGATCGGAGAGTTCGACGGTGGCTTTCGGCGTCCGCGCGTA 413  
Db 1471 CCGTCCGCGGTAGCGGACTTGGCAAGACCGCGTGGCTGTACAGGCTTCCAG 1530  
QY 414 CGCACCACCGCTTGCCTTCGCCCTTGGCCCTTGGAGGATTCAGCAGGGGAGCAGTCCGTGG 473  
Db 1531 CTGGCCAGCGGATCTCTGGCTTCGGCGGTCTGCGGTACAGCAGCTCCGCGTG 1590  
QY 474 GATGCTTCGAGTC 488  
Db 1591 CTCGCCGTGTTGCC 1605

## RESULT 31

US-09-252-991A-13657/C  
; Sequence 13657, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13657  
; LENGTH: 2175  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13657

Query Match 9.4%; Score 47; DB 4; Length 2175;  
Best Local Similarity 45.3%; Pred. No. 0.15;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;  
QY 114 CGGCGGACACCGCGGGTCACTGTCAGTGTACCGGGGAGCAGTACAGGATCCACTG 173  
Db 675 CGCAGGACACCGTCCGAGGAGATCCAGAGGCCAGCGCTGGCGAGGATCCAGGGT 516  
QY 174 TCCGCCAGCGCGGAACTCTCTGCTTCCGCAATGATCTCGTGGCGGTGTTCCAGGC 233  
Db 615 GTCGACGCGCGGTTCGGGACCAACCGCGTCTGAGACCTGATGGGAAACAGTCCGAGGC 556  
QY 234 GAAGAGCAGCGGTAGTCCACCGCGTCCGGCGGTGAACGCGTCCGGGTCGCGACCGGGAT 293  
Db 555 GACCGCCAGGAGCTTGGCCACCAACCGCGCGGTGACCTCGCGTAGGAACGCAAG 496  
QY 294 GTGCTGCGGGGTGAGCGGCCCTTGTGGCCCGCGTGTGTCGACACCGAGAGAC 353  
Db 495 CCGCTTGGCGGCGGAGAACAGTGTTCGGCAAGGCGTGGCTAGTCCGCCACCG 436  
QY 354 CAGGTCGAGCAGATCGGAGAGTTCGTCAGGTGGCGCTTTCGGCGTCCGCGCGTA 413  
Db 435 CCGTCCGCGGTAGCGGACTTGGGAGACCGCGCTGCGCTGTACAGGCTTCCAG 376  
QY 414 CGCACCACCGCTTGCCTTCGACCTTGGAGGATTCAGCAGGGGAGCAGTCCGTGG 473  
Db 375 CTGGCCAGCGGATCTCTGGTCTCCGCGCTCTCGGTACAGCAGCTCCGCGTG 316  
QY 474 GATGCTTCGAGTC 488  
Db 315 CTCGCCGTGTTGCC 301

## RESULT 32

US-09-252-991A-13575/C  
; Sequence 13575, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13575  
; LENGTH: 2292  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13575

Query Match 9.4%; Score 47; DB 4; Length 2292;  
Best Local Similarity 45.3%; Pred. No. 0.15;  
Matches 170; Conservative 0; Mismatches 205; Indels 0; Gaps 0;  
QY 114 CGGCGGACACCGCGGGTCACTGTCAGTGTACCGGGGAGCAGTACAGGATCCACTG 173  
Db 646 CCGCAGGACACCGTCCGAGGAGATCCGCGGCGCACCGCTGGCGAGGATCCACAGGGT 587  
QY 174 TCCGCCAGCGCGGGAATCTCTCTTCCGCAATGATCTCGTGGCGGTGTTCCAGGC 233  
Db 586 GTCGAAACCGCGTTCGGGACCAACCGCGTCTGAGACCTGATGCGGAACAGTCCCGAGGC 527  
QY 234 GAAGAGCAGCGGTAGTCCACCGCGTCCGGCGGTGAACGCGTCCGGGTCGCGACCGGGAT 293  
Db 526 GACCGCAGGAGCTTGGCCACCAACCGCGCGGTGACCTCGCGTAGGAACCGCAGAG 467  
QY 294 GTGCTGCGGGGTGAGCGGCCCTTGTGGCGGGCTGTGTCGACACCGAGAGAC 353  
Db 466 CCGCTTGGCGGCGGAGAACACAGTGTTCGGGCAAGCGCTGGCGTAGTCCGCGACCG 407  
QY 354 CAGGTCGAGCAGTCCGCGGAGTTCGTCAAGTGTTCGCGCGCTCTTCCCGCTCGCGCGTA 413  
Db 406 CCGCTCCGCGGTAGCGGACCTTGGCGAAGACCGCGCTGCTGACAGCGTTCAG 347  
QY 414 CGCACCACCGCTTGGCGCTTCGGCGTTGAGCGAGTTCAGCAGGGCGAGGTCGTTCCAG 473  
Db 346 CTCGCCAGCGGATCTCTGGCTTCGCGCGTCTGCGGTACCGAGCAGCTCCGCGTG 287  
QY 474 GATGCTTCGAGTC 488  
Db 286 CTCGCCGTGTTGCC 272

## RESULT 33

US-09-252-991A-9022  
; Sequence 9022, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9022  
; LENGTH: 2571

```
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9022

Query Match      9.4%; Score 46.8; DB 4; Length 2571;
Best Local Similarity 47.6%; Pred. No. 0.16;
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 208 ATGATCTCTCGCGGTGTTCCAGGCGAAGACGCGTAGTCCACCGGTCGGGCGTG 267
Db 971 ATGCTGACGCGGAGTAGGGGATGGCGATCGGTGTTCGGCATCTTGTCCGCGAGCTG 1030
QY 268 AACCGCTCCGGGGTCCGACCGGATGTCGTGCGGGGGTGAGCGGCGCTCTTGGCC 327
Db 1031 GAGCCTTTCTCGCGGAGACGCCGATGACCTGGACGCGACGTTCTCGATGAGGATGTAG 1090
QY 328 GCGTCTGTGTCACACCCAGGAGACGAGTCCGAGCCGATGCCGAGAAAGTTTCGTACG 387
Db 1091 CGGCGGATCGGGTTGGGGCTGCCGAACAGCTTCTTGGCGCACCTTGTAGCCGATCACC 1150
QY 388 GTGCGCTCTTCGCGGTGCGGCGCTAGCCACACCGCTTGCCTCGGCGCTTGAGCGAG 447
Db 1151 ACCGTGTGGGGCGCTTCTGTCGCTCGGTGAGTAGTGCCTCGGCGACCGCCAG 1210
QY 448 TTCAGCAGGCGGACGAGTGGTCCGATGCCCTCGACGTCCGCGCGGAA 497
Db 1211 TTGAGGATTTCCGGGAAGTCGTCTTGTTCGCGCGACGTAGCGTGGTA 1260

RESULT 34
US-09-252-991A-8765/c
; Sequence 8765; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8765
; LENGTH: 3150
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8765

Query Match      9.4%; Score 46.8; DB 4; Length 3150;
Best Local Similarity 47.6%; Pred. No. 0.16;
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 208 ATGATCTCTCGCGGTGTTCCAGGCGAAGACGCGTAGTCCACCGGTCGGGCGTG 267
Db 2588 ATGCTGACGCGGAGTAGGGGATGCGATCGGTGTTCGGCATCTTGTCCGCGAGCTG 2529
QY 268 AACCGCTCCGGGGTCCGACCGGATGTCGTGCGGGGGTGAGCGGCGCTCTTGGCC 327
Db 2528 GAGCCTTTCTCGCGGAGCACGCCGATGACCTGGAAACGCGACGTTCTCGATGAGGATGTAG 2469
QY 328 GCGTCTGTGTCACACCCAGGAGACGAGTCCGAGCCGATGCCGAGAAAGTTTCGTACG 387
Db 2468 CGGCGGATCGGGTTGGGGCTGCCGAACAGCTTCTTGGCGCACCTTGTAGCCGATCACC 2409
QY 388 GTGCGCTCTTCGCGGTGCGGCGCTAGCCACACCGCTTGCCTCGGCGCTTGAGCGAG 447
Db 2408 ACCGTGTGGGGCGCTTCTGTCGCTCGGTGAGTAGTGCCTCGGCGACCGGCGAG 2349
QY 448 TTCAGCAGGCGGACGAGTGGTCCGATGCCCTCGACGTCCGCGCGGAA 497
```

```
Db 2348 TTGAGGATTTCCGGGAAGTCGTGTGTTGCCGCCGAGTAGGCGTGTA 2299

RESULT 35
US-09-252-991A-8986/c
; Sequence 8986; Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8986
; LENGTH: 3450
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8986

Query Match      9.4%; Score 46.8; DB 4; Length 3450;
Best Local Similarity 47.6%; Pred. No. 0.16;
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 208 ATGATCTCTCGCGGTGTTCCAGGCGAAGACGCGTAGTCCACCGGTCGGGCGTG 267
Db 1458 ATGCTGCGGAGCGGATAGGGGATGCGGATGTCGCGCATCTTGTCCCGGAGCTG 1399
QY 268 AACCGCTCCGGGGTCCGACCGGATGTCGTGCGGGGGTGAGCGGCGCTCTTGGCC 327
Db 1398 GAGCCTTTCTCGCGGAGCACGCCGATGACCTGGAAACGCGACGTTCTCGATGAGGATGTAG 1339
QY 328 GCGTCTGTGTCACACCCAGGAGACGAGTCCGAGCCGATGCCGAGAAAGTTTCGTACG 387
Db 1338 CGGCGGATCGGGTTGGCGCTCCGCAACAGCTTCTTGGCGCACCTTGTAGCCGATCACC 1279
QY 388 GTGCGCTCTTCTCGCGGTGCGGCGGTACGCCACACCGCTTGCCTCGGCGCTTGAGCGAG 447
Db 1278 ACCGTGTGGGGCGCTCTTCTGTCGCGCTCGGTGAGTAGTGCCTCGGCGACCGGCGAG 1219
QY 448 TTCAGCAGGCGGAGCAGTCCGTGCGGATGCCCTCGACGTCCGCGGCGAA 497
Db 1218 TTGAGGATTTCCGGGAAGTCGTGTGTTGCCGCCGAGTAGGCGTGTA 1169

RESULT 36
US-07-642-734C-3/c
; Sequence 3; Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: McAlpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; CITY: Park Rd
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```







	Query Match	9.4%;	Score 46.8;	DB 3;	Length 20235;
	Best Local Similarity	46.4%;	Pred. No. 0.17;		
	Matches 153;	Conservative 0;	Mismatches 177;	Indels 0;	Gaps 0;
QY	109	GC CGGGCGGGCGGACACCGGGGGTCAAGTCTGTCACGTGTACGCGGGGACGCTACAGGATC	168		
Db	12569	GC GGGGTGGACGACGTCGACCCGGTCAGGGGTGCGGCGCCGCTGACGAGGTCTGGTCAAC	12510		
QY	169	CATGTCCGCGCCACCGCGGGAACTCCTGTCTCTTCGCCATGATCTCGTCGGCGTGTTC	228		
Db	12509	GACCAAGTCAGCTGCGGGGCCACGCGCGCTCGCAATCGCGGATCGAGTCGGCGAATACC	12450		
QY	229	CAGCGAAGACGACGCGTAGTGTCCACCGGTGCGGGGTGAACCGCTCCGGGTGCGCAC	288		

### Query Match

9.48: score 46.8; DB 2: Length 43280;

Best Local Similarity 50.7%; Pred. No. 0.17;  
Matches 139; Conservative 0; Mismatches 132; Indels 3; Gaps 1;  
QY 227 TCCAGGCGAAGAGCAGCGGTAGTCCACCGGTGCGGCGTGAACGGTCCGGGGTGGCA 286  
Db 18890 TCTGGGCGGTAGAGCGCGTCTGTCAGGACGCCCGCGTGTGAGGACCGCGTACGG 18831  
QY 287 CCGGGATGTGCGTCCCGGGGTGAGCGCGCCCTGCTTTGGCCGGCGTGGTTCGCACACCC 346  
Db 18830 GGTGGCGCGGGTACCGGTTCGAGCAGGAGGCGCGCGCGGTGCGGCGTGC 18771  
QY 347 AGGAGACAGGTCCGGACCGATCCGAGAGTTCTGTCAGGTGGCGCTTTCCGCGTCG 406  
Db 18770 AGCGGCGACGTCCG---CGTGGCGCCCAAGGCGGCCATTTCCGCGACAGTTCCGCGG 18714  
QY 407 CGCGTACGCCACACCGCGTTCGCCCTTCGCGCTTTCAGCGAGTTTCAGCAGGCGGAGCAGT 466  
Db 18713 CGCGGGGCGTGGCGCGCTTCGGTGTTCAGGAGCAGGCGCGCGGCGACCGTCCGCGG 18654  
QY 467 CGTGGCGATGCCCTTCAGCTCGCGCGGCGAACCT 500  
Db 18653 TGACAGATGCGGCGCGCTTCGCGCGGAGCT 18620

## RESULT 39

US-08-576-626A-2/c  
; Sequence 2, Application US/08576626A  
; Patent No. 5998194  
; GENERAL INFORMATION:  
; APPLICANT: Summers, R.G.  
; APPLICANT: Katz, L.  
; APPLICANT: Donadio, S.  
; APPLICANT: Staver, M.J.  
; TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR  
; TITLE OF INVENTION: BIOSYNTHESIS GENES  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA: US/08/576.626A  
; FILING DATE: 21-DEC-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Diane Casuto  
; REGISTRATION NUMBER: P-40,943  
; REFERENCE/DOCKET NUMBER: 5857.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (847) 938-3137  
; TELEFAX: (847) 938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8051 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-576-626A-2

Query Match 9.3%; Score 46.6; DB 2; Length 8051;  
Best Local Similarity 49.8%; Pred. No. 0.18;

Matches 118; Conservative 0; Mismatches 119; Indels 0; Gaps 0;  
QY 1 GGTATCCCGACCGTGTCCCGGAACAACGAGTCGAGATACGGGAGGGAACACCCCGGT 60  
Db 7517 GATGCCCAACCGCGCCACGAAACGCTCTCTCGAGCGGAGACACGAAAGGCCCCGCC 7458  
QY 61 AGTCGGGTAGACGCTGGCGGCGAAGGCTACGGCCCTTCGACGGTCAAGCGGCGGGCGG 120  
Db 7457 GGTCTGTCTGGGTTCGGGGTGAACTCGAAACCGCCCGCCAGCTCACTACTC 7398  
QY 121 ACACCGCGGGTACGCTCGTCAGCTGACGCGGGGAGTACAGGATCCACTGTCCGCA 180  
Db 7397 GCAATCTCGTTCACTTCGCGGAGGAGCGGGTGTCTCCACCCCTCGCGGATG 7338  
QY 181 GCCCGCGGAACTCCTGTCTTCGCCATGATCTGTCGGGTGTTCAGGCGAAG 237  
Db 7337 TCTCTTCGAGTGTGAGCGGTTCGGGTACCGACTGCTGCGAAGCGGTTCAGGGAAG 7281

## RESULT 40

US-09-266-965-98/c  
; Sequence 98, Application US/09266965  
; Patent No. 6495348  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D  
; APPLICANT: Mao, Y  
; APPLICANT: Varoglu, M  
; APPLICANT: He, M  
; APPLICANT: Sheldon, P  
; TITLE OF INVENTION: Mitomycin biosynthetic gene cluster  
; FILE REFERENCE: 600.456U1  
; CURRENT APPLICATION NUMBER: US/09/266.965  
; CURRENT FILING DATE: 1999-03-12  
; EARLIER APPLICATION NUMBER: US 08/624,447  
; EARLIER FILING DATE: 1996-08-19  
; EARLIER APPLICATION NUMBER: PCT/US94/11279  
; EARLIER FILING DATE: 1994-10-06  
; EARLIER APPLICATION NUMBER: US 08/133,963  
; EARLIER FILING DATE: 1993-10-07  
; NUMBER OF SEQ ID NOS: 145  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 98  
; LENGTH: 819  
; TYPE: DNA  
; ORGANISM: Streptomyces lavendulae  
US-09-266-965-98

Query Match 9.3%; Score 46.4; DB 4; Length 819;  
Best Local Similarity 46.6%; Pred. No. 0.19;  
Matches 149; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 176 CGCAGCCCGCGGAACTCTCTCTTCGCCATGATCTGTCGGCGTGTTCAGGCGA 235  
Db 334 CGCGGCGCGCGGACCTCGTTCGGGCGGCTTCTCGACAGCGCGGAGAACCGCGCG 275  
QY 236 AGAGCAGCGGTAGTCCACCGGTTCGCGGTGAAACCGTTCGGGTGCGCACCGGGATGT 295  
Db 274 CGTGGGAGTTTCGCGACCGCGCGTGTGAGGCGTTCGCTCCCGCGCGCGGT 215  
QY 296 GCGTGC CGGGGTGAGCCCGCCCTCTCTTGGCGCGGTCTGTCGACACACCGAGGAGCA 355  
Db 214 GTGCGCGAGGAAGCGCGGCCCCACAGACGTCGTCGAGGAGAACCGCGCGGCG 155  
QY 356 GGTCCGACCGATGCGCAGAGTTCTGTCAGGTGGCGCTCTTCGCGTCCGCGGTACG 415  
Db 154 CGAGGCGCGCGCGCCCGCGCGGTTCGCGGCGCGCGCGCGCGCGCGCGCGGTTCGCCGA 95  
QY 416 CCACCAACCGCTTTCGCGCTTCGCGCTTCGAGCGATTTCAGCAGGCGCGGAGGTCTCGT 475  
Db 94 CCAAGCGCAGCGGAGAGCGCGCTCATCTCGCCTGTTCAGCGCGATCAGTTGATCACCGG 35  
QY 476 TGCCTTCGACGTCCGCGCGG 495

Db 34 CGTCGCCCGCGTGGTGTG 15

## RESULT 41

US-09-252-991A-10644/c  
; Sequence 10644, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10644  
; LENGTH: 1029  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10644

Query Match 9.3%; Score 46.4; DB 4; Length 1029;  
Best Local Similarity 45.1%; Pred. No. 0.2;  
Matches 173; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY	116	GGCGGACACCGCGGCTCGTACGTCAGCGGGGAGTACAGATCCACTGTC	175
Db	559	GCGATCCCGCGGTGCGGAGGTAGATCTCCCGGATACAGGCGCAGCGCTGCG	500
QY	176	CGCCAGCCCGCGGAACTCTGCTTCGCGATGATCTGCGCGGTGTTCCAGGGGA	235
Db	499	CGGATCGTAGCGACGTCCAGCGCTGTCGCTGTTTATCTCGACCTGGCTGCTCT	440
QY	236	AGAGCAGCGGTAGTCCACCGGTGCGGCGTGAACCGGTGCGGCGTGGTCCAGGGGA	295
Db	439	CGAGCACCAGGTGCGCGCTCGCGACGCGAGTGAAGTGAAGTGAAGTGAAGTGAAG	380
QY	296	GCGTCCCGCGGTGAGCCGCGCTGCTTGGCCGCGGTGCTGTCGACACCCAGGAGACA	355
Db	379	GCCAGACTTGGCGCGACATCCCGACAGCGGTGCGCGCGCGGAGCAGCGGGGA	320
QY	356	GGTCCGACCGTATCCCGAGAGTGGTCACTGCGCGCTTTCGCGCGTGGCGCGTACG	415
Db	319	GGCGCGCAGGACCGCGCGCGCTGCGCGAGCGCTCGAGCGCAGCGCGGCGCA	260
QY	416	CCACCAACCGCTTGGCGCTGAGCGAGTTCAGCAGGCGGAGCGGTGCGTCCGGA	475
Db	259	GCGCGCGCGCAGCGAACCAGATATTCGCCAGCAGCTCGCGCGCGCGCGCGCT	200
QY	476	TGCGCTCGAGTGGCGGCGCAACC	499
Db	199	GGTGCTCGGCGCTGCGCGAGCGCC	176

## RESULT 42

US-09-252-991A-10533/c  
; Sequence 10533, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10533  
; LENGTH: 1341  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10533

Query Match 9.3%; Score 46.4; DB 4; Length 1341;  
Best Local Similarity 45.1%; Pred. No. 0.2;  
Matches 173; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

QY	116	GGCGGACACCGCGGCTCGTACGTCAGCGGGGAGTACAGATCCACTGTC	175
Db	545	GCCGATCGCGCGGTGCGGAGGTAGTCTCCGATACAGCGCGCAGGGCGTGGC	486
QY	176	CGCCAGCCCGCGGAACTCTGCTTCGCGATGATCTCGCGGTGTTCCAGCGGA	235
Db	485	CGGATCGTAGCGACGTCCAGCGCGCTGCGGTGTTTATCTCGACCTGGCTGCTCT	426
QY	236	AGAGCAGCGGTAGTCCACCGGTGCGGCGTGAACCGGTGCGGCGTGGTCCAGGGGA	295
Db	425	CCAGCACCAGTCCGCGCGCTCGCGCAGCGAGTGAAGTGAAGTGAAGTGAAGTGAAG	366
QY	296	GCGTCCCGCGGTGAGCCGCGCTGCTTGGCCGCGGTGCTGTCGACACCCAGGAGACA	355
Db	365	GCCAGACTTGGCGCGACATCCCGACAGCGGTGCGCGCGGAGCAGCGGGGA	306
QY	356	GGTCCGACCGTATCCCGAGAGTGGTCACTGCGCGCTTTCGCGCGTGGCGCGTACG	415
Db	305	GCGCGCGCAGGACCGCGCGCGCTGCGCGAGCGCTCGAGCGCAGCGCGGCGCA	246
QY	416	CCACCAACCGCTTGGCGCTGAGCGAGTTCAGCAGGCGGAGCGGTGCGTCCGGA	475
Db	245	GCGCGCGCGCAGCGAACCAGATATTCGCCAGCAGCTCGCGCGCGCGCGCGCT	186
QY	476	TGCGCTCGAGTGGCGGCGCAACC	499
Db	185	GGTCTCGGCGCTGCGCGAGCGCC	162

## RESULT 43

US-09-252-991A-10409/c  
; Sequence 10409, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10409  
; LENGTH: 1398  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10409

Query Match 9.3%; Score 46.4; DB 4; Length 1398;  
Best Local Similarity 48.5%; Pred. No. 0.2;  
Matches 158; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

QY	47	GAACACCCCCCGGTAGTCCGCGGTAGCGGTGCGGCGGAGCGGTACGCGCTTCGCGGT	106
Db	522	GATGATCACCGGTAGCGGACCTCGCGGCGATCGCCAGGCGGCTTCTTTCATCTCCGG	463
QY	107	CAGCGGCGGCGGACACCGG--CGGTGAGTCTGTCAGGTGACGCGGCGGAGTACAG	164
Db	462	CAGCGGCGGCTCGGAGGCCCGCGGAGCGGTGCGGAGCGCGCGCGCTTCTATGCGCTTGGC	403

165 GATCCACTGTCGCGCAGCCGCGGAACTTCTGCTCTCTTCCGCCATGATCTCTGTCGCGGTG 224  
Db 402 CGAGACTTGTGTCGCCCATCAGGCGGATCACCTCGGCGGTGCGGCCCGCAAGGTGAAGCC 343  
Qy 225 GTTCCAGGCGAAGAGCAGCAGCGCTAGTCCACCGCGTGGGCGTGAACCGCTCCGGGGTGG 284  
Db 342 GGAGCGTTTCGATCTGCTCGGCGAAGTCGGCGTTCTCGGCGAGGAAGCCGTAGCCGGGTG 283  
Qy 285 CACCGGGATGTGGTCCCGGGGTGAGCCGGCCCTGCTTGGCGCGCGTGTGTCGCACAC 344  
Db 282 GATCGCGGTGGCCCGCGGTGACCTCGGCGCGCGGATGATCGCCGGGATCTGCAGGTACGA 223  
Qy 345 CCAGGAGACAGTCCCGGACCGATGC 370  
Db 222 CTGGGTGGCGGGCGGACCGATGC 197

RESULT 44  
US-09-252-991A-10752  
; Sequence 10752, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10752  
; LENGTH: 1446  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10752

Query Match 9.3%; Score 46.4; DB 4; Length 1446;  
Best Local Similarity 48.5%; Pred. No. 0.2;  
Matches 158; Conservative 0; Mismatches 166; Indels 2; Gaps 1;

Qy 47 GACACCCCCCGGTAGTCCGGGTAGACGGTGGGCGGAGCGGTACCGCGCTTCGACGGT 106  
Db 952 GATGATCACCGGGTAGCGACCTCGCGGGCGATCGCCAGGGCGGTTTCTTCATCTCCGG 1011  
Qy 107 CAGCGGCGCGCGGACACCGG--CGGGTCAGTCTGTCACGTGTACGCGGGGACGTACAG 164  
Db 1012 CAGCGGCGCGTGGAGCCCGGACGGTGGGACGCCCGCGCGCTTCATGCGCTCTTGGC 1071  
Qy 165 GATCCATGTGTCGCGCAGCGCGCGGGAATCTCTGTCTCTTCGCCCATGATCTGTCGCGGTG 224  
Db 1072 CGAGACCTTGTTCGCCCATCAGCGCGGATCACCTCGCGCGGTTCGGGCGCGAGGTGAAGCC 1131  
Qy 225 GTTCCAGCGAAGACGACGCGGTAGTCCACCGCGTCCGGGCGTGAACCGCTCCGGGGTGG 284  
Db 1132 GGACGGTTCGATCTGCTTCGCGGAATCGGGTTCCTCGGCGAGGAGCGGTAGCCGGGTG 1191  
Qy 285 CACCGGATGTGCGTGGCGGGGTGAGCGGCGCTCTGTTGGCGCGCGCTGCTGTGTCGCACAC 344  
Db 1192 GATCGCGGTGGCGCGGTGACCTCGGCGCGGCGATGATCGCCGGGATCTGCAGGTACGA 1251  
Qy 345 CCAGGAGACAGGTCCGACCGATGC 370  
Db 1252 CTGGGTGGCGGGCGGACCGATGC 1277

RESULT 45  
US-09-266-965-9/c  
; Sequence 9, Application US/09266965  
; Patent No. 6495348

```
; NUMBER OF SEQ ID NOS: 145
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 76
; LENGTH: 53500
; TYPE: DNA
; ORGANISM: Streptomyces lavendulae
US-09-266-965-76

Query Match          9.3%; Score 46.4; DB 4; Length 53500;
Best Local Similarity 46.8%; Pred. No. 0.21;
Matches 149; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY      176   CGCACGCCGGCGGAACCTCCTGTCCTTCCGCATCATCTCGTCGCGGTGTTCAGAGCA    235
        |||         |       |             |              |            |
Db      19824 CGCGGCGCGCGGAGACTTGCTGGGCGCGCTTCTCNACAGCGCGAGAACCGCGCG    19883

QY      236   AGAGCAGCGGTAGTTCCAICGGTCTGGGCGTGAAACGCTCGGGGTGCGCACCGGGATTG    295
        ||||         |           |     |||               |      |||
Db      19884 CGETGGGGTAGTTCCGCACAGCGACCCTGGGTGAAGCGCTCCGGTTCGCGCGCGCGCGT    19943

QY      296   GCCTGCGCGGGGTGAGCGCGCCCTGTTGCGCGGGCTGTCGACACCCCAGGAGAACA    355
        |||         |           |           |             |      |||
Db      19944 GTGCGGAGAGAACCGGGCCCCAACCAGAGCTCGTCTCGAGGAAGACCAAGCCGGSG    20003

QY      356   GGTCGCGAACCGATGCCGCGAGAAAGTTGCTCACGGTGGCGCTCTTCGCGTFCGCGCCGTNAG    415
        ||||         |           |           |             |      |||
Db      20004 CGAGGACCGGGCGCGCCCGCGGCTTCGCGGCGCGCGCCCGCCCGCAGCGTCCCGCA    20063

QY      416   CCACCAACCGCTTGCCTTCGGCCCTTCAGCGAGTTACAGAGGCGGACGAGTCCGGTCGGA    475
        ||||         |           |           |             |      |||
Db      20064 CCAAGCGACGCGGAGAAAGCCCGCTCATCTCGCCTGTACAGCGGATCAGTTGATCACCGG    20123

QY      476   TGCCCTCGACGTCGGCGGG    495

Db      20124 CGTCGCCCCGCTGTTGTG    20143
```

```

RESULT 47
US-09-252-991A-122/c
; Sequence 122, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 122
; LENGTH: 822
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-122

```

Query Match	9.2%;	Score 46.2;	DB 4;	Length 822;
Best Local Similarity	46.2%;	Pred. No. 0.21;		
Matches 153; Conservative 0;	Mismatches 178;	Indels 0;	Gaps 0;	
Qy	138	CGTCACGTGTACGGGGGGACGTACAGGATCACTGTCCGCGACGCGCGGAACTCTCGT	197	
Db	712	CGCCTCGCGTTGCGCGCGCGAGCGCGCCGCGAGGTCGCGCGGGCGCCGAGGATATC	653	
Qy	198	CTCCTTCGCCATGATCTCGTCGGCGTGGTTTCAGCGGAAGACAGCGGTTAGTCACCGC	257	
Db	652	GGCCAGGGCCCTCGCCCAAGCAGGTGCGACCCGACCCAGTTCCTCCGTCGACGGCACTC	593	
Qy	258	GTCCGGGGGTGAACCGGTCCGGGGTGCGCACCGGATGTGCGTGGCGGGGGTGAGCCGGCC	317	

592 GACGGTCTCACCGGAAGGGTCTTCAGAGCAGGCGCAGCGAGCGTAGCTGTCGATGTC 533

318 CTCCTTGGCGGGCGGTCTGTCGACACCCAGGAGACCAAGTTCGGAGCCGATGCCGAGAA 377

532 GTAGGCGCATCGGGCGGACCTTCGCCGCCAACGCCTCCAGTTCTCGACGCTGGCAGGGT 473

378 GTTTCGTACAGGTGGCGGCTCTTTCGCCCTCGCGCGGTACGCCACCAACCGGCTTGCCTTCGGC 437

472 GATCCGCGCGCGCGGCGCTTGCCTCATGGCGTGAGACCATCACCCCGGATCGTGAAGGC 413

438 CTTGAGCGAGTTCAGCAGGGCGAGCAGTTCG 468

412 GATCAGCGCGTTGGCCTGTAGCCGTGGCGG 382

RESULT 48

US-09-252-991A-148

; Sequence 148, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 148

; LENGTH: 936

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-148

RESULT 49  
US-09-252-991A-129/c  
; Sequence 129, Application US/09252991A  
; Patent NO. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 18:07:09 ; Search time 339.02 Seconds  
(without alignments)  
6516.033 Million cell updates/sec

Title: US-09-758-759-1\_COPY\_109000\_109519  
Perfect score: 520  
Sequence: 1 ggtgcagcagcagcagcgg.....ggccagcgccgcgggatcc 520

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : N\_Geneseq\_29Jan04.\*

- 1: N\_Geneseq1980s.\*
- 2: N\_Geneseq1990s.\*
- 3: N\_Geneseq2000s.\*
- 4: N\_Geneseq2001as.\*
- 5: N\_Geneseq2001bs.\*
- 6: N\_Geneseq2002as.\*
- 7: N\_Geneseq2003as.\*
- 8: N\_Geneseq2003bs.\*
- 9: N\_Geneseq2003cs.\*
- 10: N\_Geneseq2004s.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	520	100.0	109519	5	AAS08693
2	86.2	16.6	1954	2	AQ12604
3	58.2	11.2	114955	2	AX53491
4	55.2	10.6	2514	7	ACA40308
5	55.2	10.6	110000	4	AA199682_03
6	53.2	10.2	6941	9	ADC30320
7	52	10.0	4638	7	ACF34536
8	51.6	9.9	114955	2	AX53491
9	51.2	9.8	89421	6	AA140781
10	50.6	9.7	975	7	ABZ66676
11	50.6	9.7	45055	7	ABZ66680
12	50.4	9.7	1209	7	AA161198
13	50.4	9.7	2684	7	ABZ77318
14	50.4	9.7	82746	7	AA161224
15	50	9.6	2811	6	AA161221
16	49.6	9.5	1554	7	AA161221
17	49.2	9.5	11905	7	AA161225
18	49.2	9.5	1575	7	ACA37544
19	49	9.4	2000	7	ADA71938
20	49	9.4	5760	6	ABZ78661
21	49	9.4	24081	7	AA161223
22	49	9.4	52101	7	AA161223
23	49	9.4	109519	5	AAS08693

C	24	49	9.4	113193	7	AD54645	Aad54645 Streptomy
C	25	48.8	9.4	2208	7	ACA26655	ACA26655 Prokaryot
C	26	48.6	9.3	795	7	AA161219	AA161219 Actinosyn
C	27	48.6	9.3	1797	7	ACA40900	ACA40900 Prokaryot
C	28	48.6	9.3	11905	7	AA161225	AA161225 Actinosyn
C	29	48.6	9.3	110000	4	AA199683_03	Continuation (4 of
C	30	48.6	9.3	110000	4	AA199683_39	Continuation (40 o
C	31	48.4	9.3	1689	7	ACA27331	ACA27331 Prokaryot
C	32	48.4	9.3	3659	7	ACA38077	ACA38077 Prokaryot
C	33	48.2	9.3	90800	6	ABQ78872	ABQ78872 S. roseos
C	34	48	9.2	536	9	ADB68842	ADB68842 Minority
C	35	48	9.2	3957	5	AAA09686	AAA09686 HSV-2 inn
C	36	48	9.2	154746	6	AA161225	AA161225 Human her
C	37	48	9.2	154746	6	AA161225	AA161225 Human her
C	38	47.6	9.2	1679	7	ABZ78139	ABZ78139 Human can
C	39	47.6	9.2	4523	9	AD560227	AD560227 Human gen
C	40	47.6	9.2	4523	9	AD560229	AD560229 Human gen
C	41	47.6	9.2	4523	9	AD560231	AD560231 Human gen
C	42	47.6	9.2	4523	9	AD560233	AD560233 Human gen
C	43	47.6	9.2	4523	9	AD560233	AD560233 Human gen
C	44	47.4	9.1	88421	6	AA140781	AA140781 88421nt g
C	45	47.2	9.1	2307	6	ABK32842	ABK32842 DNA encod
C	46	47.2	9.1	5571	7	ABX72218	ABX72218 Human NOV
C	47	47	9.0	1511	7	ACA42277	ACA42277 Prokaryot
C	48	47	9.0	1835	3	AA130798	AA130798 DNA encod
C	49	47	9.0	5997	6	AA168598	AA168598 A. cryptu
C	50	47	9.0	110000	4	AA199682_39	Continuation (40 o

ALIGNMENTS

RESULT 1

AAS08693

ID AAS08693 standard; DNA; 109519 BP.

XX AAS08693;

XX AAS08693;

DT 11-SEP-2003 (revised)

DT 26-SEP-2001 (first entry)

XX

Micromonospora DNA encoding biosynthetic enzymes for Everninomycin.

XX Everninomycin; antibiotic; bottle-neck gene; orthomycin; fermentation;

XX ds.

XX Micromonospora sp. ATCC 39149.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

FT	CDS	/tag= k	FT	RBS	complement(31885. .31888)
FT		6232. .7275	FT		/tag= ak
FT		/tag= j	FT	CDS	complement(31941. .32882)
FT	CDS	/product= "EvdP"	FT		/tag= al
FT		7272. .8327	FT		/product= "EvrL"
FT		/tag= l	FT	CDS	complement(33167. .34405)
FT		/product= "EvdG"	FT		/tag= am
FT	RBS	8333. .8336	FT		/product= "EvrM"
FT		/tag= n	FT	RBS	complement(34414. .34418)
FT	CDS	8342. .9364	FT		/tag= an
FT		/tag= m	FT	CDS	complement(34449. .35210)
FT		/product= "EvdH"	FT		/tag= ao
FT	CDS	complement(9463. .10224)	FT		/product= "EvrN"
FT		/tag= o	FT	RBS	complement(35219. .35221)
FT		/product= "EvdI"	FT		/tag= ap
FT	RBS	complement(10232. .10235)	FT	CDS	complement(35294. .36238)
FT		/tag= p	FT		/tag= aq
FT	CDS	10424. .11176	FT		/product= "EvrO"
FT		/tag= q	FT	CDS	complement(36235. .36963)
FT		/product= "EvdJ"	FT		/tag= ar
FT	CDS	12027. .12455	FT		/product= "EvrP"
FT		/tag= r	FT	CDS	complement(36998. .38026)
FT		/product= "EvdK"	FT		/tag= as
FT		/partial	FT		/product= "EvrQ"
FT		/note= "No start codon"	FT	CDS	complement(38072. .38566)
FT	CDS	complement(12108. .13022)	FT		/tag= at
FT		/tag= s	FT		/product= "EvrR"
FT	RBS	/product= "EvdL"	FT	CDS	complement(38892. .40163)
FT		/tag= t	FT		/tag= au
FT	CDS	complement(13027. .13030)	FT		/product= "EvrS"
FT		/tag= u	FT	CDS	complement(40216. .40890)
FT		/product= "EvrA"	FT		/tag= av
FT	RBS	complement(15369. .15373)	FT	CDS	/product= "EvrT"
FT		/tag= v	FT		complement(40887. .41576)
FT	CDS	complement(15380. .16414)	FT		/tag= ax
FT		/tag= w	FT		/product= "EvrU"
FT		/product= "EvrB"	FT	RBS	complement(40899. .40902)
FT	CDS	complement(16419. .17873)	FT		/tag= aw
FT		/tag= x	FT	CDS	complement(41679. .42707)
FT		/product= "EvrC"	FT		/tag= ay
FT	CDS	complement(17870. .18934)	FT	RBS	/product= "EvrV"
FT		/tag= y	FT		complement(42714. .42717)
FT		/product= "EvrD"	FT	CDS	/tag= az
FT	CDS	19374. .20906	FT		complement(42810. .43799)
FT		/tag= z	FT		/tag= ba
FT		/product= "EvrE"	FT	CDS	/product= "EvrW"
FT	RBS	21056. .22542	FT		complement(43799. .44866)
FT		/tag= ab	FT		/tag= bc
FT	CDS	21064. .22542	FT		/product= "EvrX"
FT		/tag= aa	FT	RBS	complement(43807. .43811)
FT		/product= "EvrF"	FT	CDS	/tag= bb
FT		22736. .22740	FT		complement(45014. .45760)
FT	RBS	22748. .24172	FT		/tag= bd
FT	CDS	/tag= ac	FT		/product= "EvrY"
FT		/product= "EvrG"	FT	RBS	complement(45767. .45770)
FT		complement(24177. .25223)	FT		/tag= be
FT	CDS	/tag= ae	FT	RBS	/tag= bg
FT		/product= "EvrH"	FT		complement(45952. .45956)
FT	RBS	complement(25230. .25233)	FT	CDS	complement(45962. .46714)
FT		/tag= af	FT		/tag= bf
FT	CDS	25550. .26626	FT		/product= "EvrZ"
FT		/tag= ag	FT	CDS	complement(47156. .49234)
FT		/product= "EvrI"	FT		/tag= bh
FT	RBS	26672. .26676	FT		/product= "EvaA"
FT		/tag= ai	FT	CDS	51627. .52715
FT	CDS	26685. .30479	FT		/tag= bi
FT		/tag= ah	FT		/product= "EvaB"
FT		/product= "EvrJ"	FT	RBS	51629. .51622
FT	CDS	complement(30557. .31876)	FT		/tag= bj
FT		/tag= aj	FT	CDS	52889. .53557
FT		/product= "EvrK"	FT		/tag= bk
FT			FT		/product= "EvaC"
FT			FT	CDS	53554. .54207









Human; diagnostic; drug screening; forensics; gene mapping; biodiversity assessment; Parkinson's disease; Alzheimer's disease; neurodegenerative diseases; anaemia; platelet disorder; wound; burns; ulcers; osteoporosis; autoimmune disease; cancer; molecular weight marker; food supplement; antiparkinsonian; nootropic; neuroprotective; antianemic; anticoagulant; thrombolytic; vulnerary; antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic; gene therapy; chromosome 13q34; gene; ss.

Homo sapiens.

WO2003029271-A2.

10-APR-2003.

24-SEP-2002; 2002WO-US030474.

24-SEP-2001; 2001US-0324631P.

(HYSE-) HYSEQ INC.

Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T; Zhou P, Chosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G; Haley-Vicente D, Dmanac RT;

WPI; 2003-371981/35.

P-PSDB; ADC31291.

New polynucleotide and polypeptide useful for diagnosing, preventing or treating conditions such as neurodegenerative diseases, anemias, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer.

Claim 1; SEQ ID NO 402; 1185bp; English.

The invention relates to 971 novel human cDNA sequences (ADC29919-ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The invention also relates to nucleic acid sequences over 93% identical with the novel human cDNAs. The invention additionally encompasses expression vectors and host cells comprising a nucleic acid of the invention; the recombinant production of a polypeptide of the invention; an antibody against a polypeptide of the invention; a method of detecting polynucleotides or polypeptides of the invention; and methods of identifying a compound which binds to a polypeptide of the invention. The invention further discloses methods of preventing, treating or ameliorating a medical condition; kits comprising polynucleotide probes and/or monoclonal antibodies for carrying out the methods of the invention; methods for the identification of compounds that modulate the expression or activity of the polynucleotide and/or polypeptide; and 767 contig sequences corresponding to the cDNA sequences of the invention (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628-ADC33394). The nucleic acids and polypeptides of the invention are useful in diagnostics, drug screening, forensics, gene mapping, in the identification of mutations responsible for genetic disorders or other traits, for assessing biodiversity, and in producing many other types of data and products dependent on DNA and amino acid sequences. They are also used for treating diseases such as Parkinson's disease, Alzheimer's disease and other neurodegenerative diseases, anaemia, platelet disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or cancer. The nucleic acids may also be used as hybridisation probes or primers, and in the recombinant production of a protein. The polypeptides are also useful in generating antibodies, as molecular weight markers, and as food supplements. The present sequence represents a specifically claimed human cDNA sequence of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pat\\_sequences](http://ftp.wipo.int/pub/published_pat_sequences).

Sequence 6941 BP; 1361 A; 2228 C; 2015 G; 1337 T; 0 U; 0 Other;

Query Match 10.2%; Score 53.2; DB 9; Length 6941;

Best Local Similarity 46.08; Pred. No. 0.49;

Matches 216; Conservative 0; Mismatches 253; Indels 1; Gaps 1;

QY 46 ACTGCCACTCCGGCGGGGTGACCAAGGTGAGTGGAGTTGAGCAGCCAGCTCATCG 105  
 Db 1348 AGTTCCGGCGCGCAGTAGAAGAGCAATGTCGGGGTCTGTCGCCACCGACCCCATCAGG 1407  
 QY 106 CGCCTTGCGCGGATGCGGAAACACCGGGCCAGGATCAGTGCAGCAGCGGATGCGCG 165  
 Db 1408 TCCCCGGGCGCGCGCCACACCACTGCTGCACTCCCGCCAGCCAGACGGGCTGG 1467  
 QY 166 CCTCGATCTTCAGACGGTCCGGCGCGCAGCTCGATCTCGTCGCCCGGCTCCACACCAAGGGGAA 225  
 Db 1468 TCGCGCGCTCGCGCACCGACAGCTGGCGGCACCCCGCGCGGCGCAAGTGCAGTCTGT 1527  
 QY 226 ACTGGCTCGTGGCAGCGGAGCGCCCGCCAGCCGGGACAGCTCGTCAGGGTGGCGTCAACGGA 285  
 Db 1528 GCGGGTGGCAGCCCGCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1587  
 QY 286 CCTCTCGCGGGTCCGGGAGCGGAAACCGCAGCGGTTCAGATCCCTGTCAGTGCATCGGCT 345  
 Db 1588 CGCAGAGGCTGGTGTGCGGTGGGAGC-CCCTGAGCCCGCGGCGGCGGCGGCGGCGG 1646  
 QY 346 CAGTCCCGGTCTGTCCTCCCTTGGCCCTGGGAGGATAGCGGTTACAGACAGCGGACCAACG 405  
 Db 1647 CTGAGCCGCTCGCACACACCTGAGCGGCGGCTGCGGCGGCGGCGGAGCAAGGTGGCGCTG 1706  
 QY 406 CGGCGCGGGTCCGGGCGGCGGTTTCAGCCGATCCGCTCGATGACCGGCTGCGGGGTGCGG 465  
 Db 1707 CTGCGCGGAGGGGCGCGCTGCACACAGCCGCTCATGTCTATGTCGCCGTGGCGACTCG 1766  
 QY 466 GCGGTGCGCGAGATCCGTACCGCGGACCGCCTCGGCGGCGGCGGCGGCGGCGG 515  
 Db 1767 CGCGCGCGCGCCACCGAGCCCGGCTCCCTGCTGCTCCAGCAGCGGCGGCGG 1816

# RESULT 7

ACF34536  
 ID ACF34536 standard; DNA; 4638 BP.

AC ACF34536;

DT 15-OCT-2003 (first entry)

DE Gene encoding angiogenesis protein ENO171.

XX Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;  
 XX antipsoriatic; antiarteriosclerotic; cardiast; vasotropic; angiogenesis;  
 XX gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;  
 XX diabetic retinopathy; cardiovascular disease; atherosclerosis;  
 XX ischemic limb disease; coronary artery disease; gene; ds.

OS Homo sapiens.

XX WO2003027285-A1.

XX 03-APR-2003.

XX 19-SEP-2002; 2002WO-AU001282.

XX 27-SEP-2001; 2001AU-00007973.

XX 27-SEP-2001; 2001AU-00007974.

XX 11-OCT-2001; 2001AU-00008210.

XX 29-OCT-2001; 2001AU-00008532.

XX 13-NOV-2001; 2001AU-00008838.

XX 28-AUG-2002; 2002AU-00951032.

XX (BION-) BIONOMICS LTD.

XX Gamble JR, Hahn CN, Vadas MA;

XX WPI; 2003-354655/33.

XX P-PSDB; ABR64261.

PT New angiogenic genes and polypeptides, useful for diagnosing,

PT prognosticating or treating an angiogenesis-related disorder, e.g.  
 PT cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or  
 PT cardiovascular diseases.  
 XX  
 PS Claim 2; SEQ ID NO 91; 90pp; English.  
 XX  
 CC The invention relates to the isolation of novel genes (ACF34446-ACF34559)  
 CC encoding proteins (ABR64180-ABR64281) involved in the process of  
 CC angiogenesis. The nucleic acid molecules are useful in identifying and/or  
 CC obtaining full-length human genes involved in an angiogenic process. The  
 CC nucleic acid molecule, polypeptides or complexes encoded, cells or  
 CC genetically modified non-human animals derived from these are useful for  
 CC the screening of candidate pharmaceutical compounds used in treating  
 CC angiogenesis-related disorders. They are also useful for diagnosing,  
 CC prognosticating or treating an angiogenesis-related disorder, which  
 CC involves uncontrolled or enhanced angiogenesis or is a disorder in which  
 CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,  
 CC diabetic retinopathy, psoriasis or cardiovascular diseases such as  
 CC atherosclerosis), or involves inappropriately arrested or decreased  
 CC angiogenesis or is a disorder in which an expanding vasculature is of  
 CC benefit (e.g. ischemic limb disease or coronary artery disease). The  
 CC modulator of expression or activity of the polypeptide encoded by the  
 CC nucleic acid sequence is useful for manufacturing a medicament for the  
 CC treatment of an angiogenesis-related disorder. This sequence corresponds  
 CC to the gene encoding one of the novel angiogenic protein  
 XX  
 SQ Sequence 4638 BP; 1352 A; 957 C; 1008 G; 1321 T; 0 U; 0 Other;

Query Match 10.0%; Score 52; DB 7; Length 4638;  
 Best Local Similarity 47.9%; Pred. No. 0.83;  
 Matches 181; Conservative 0; Mismatches 195; Indels 2; Gaps 1;  
 QY 6 AGCAGAAACACGGTGTGCGCCCGCGCGGCTGAGCGGCTCGACAGCGAGCGGCGG 65  
 DB 141 AGCCCGCGGCACAGCGCGGCGGCTGAGCGGCTCGACAGCGAGCGGCGGCGG 200  
 QY 66 ACCAGGTGACAGTGGAGTGGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGG 125  
 DB 201 CCGGCT 260  
 QY 126 AAACACCGCGGCGAGGATCAGCTGACGACGCGGCGATGCGCGCTCTGATCTCAGCGTGGC 185  
 DB 261 AGACCGGCTCTCGCGCGCGCGCGCGCTCTGAGGAGCGCGCGCGCGCGCGAG 320  
 QY 186 GCAGCTGATCTCTGCGCGCGCTCTCCACACAGGAGAACTGCTCGTGGTGGAGCGG 245  
 DB 321 AGCGCGCGCGCTGGAGCGCG--TCGGGCGCGCGCGGAAAGTGGCTCTCGCGGCGCGG 378  
 QY 246 AGCCCGACGCGGACAGCTGCTGAGGCTGCGTCAAGGAGCTCTCGCGGCTCGGAGCG 305  
 DB 379 GAGCGCGGCTGCG 438  
 QY 306 GAAACGCGCGGCTCAGATCTCTGATCGCATCGGCTCAGTCCGCGTCTGTCCTT 365  
 DB 439 GAGCGGACGCGCGGCTCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 498  
 QY 366 GCGCTGGGAGATACGG 383  
 DB 499 CTCGCGCGCGCGCGCG 516

## RESULT 8

AA53491  
 ID AA53491 standard; DNA; 114955 BP.  
 XX  
 AC AA53491;  
 XX  
 DT 05-JUL-1999 (first entry)  
 XX Human adenosine A1 receptor antisense oligonucleotide fragment.  
 DE  
 XX Antisense oligonucleotide; multiple target; antisense treatment;  
 KW impaired respiration; inflammation; lung disease;  
 KW

KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
 KW acute asthma; allergy; asthma; impeded respiration;  
 KW respiratory distress syndrome; pain; cystic fibrosis;  
 KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
 KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
 KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
 KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
 KW prostate cancer; ss.  
 XX Synthetic.  
 OS  
 XX WO9913886-A1.  
 XX  
 PD 25-MAR-1999.  
 XX  
 PF 17-SEP-1998; 98WO-US019419.  
 XX  
 PR 17-SEP-1997; 97US-0059160P.  
 PR 09-JUN-1998; 98US-00093972.  
 XX  
 PA (UVEC-) UNIV EAST CAROLINA.  
 XX  
 PI Nyce JW;  
 XX  
 DR WPI; 1999-229400/19.  
 XX  
 PT New antisense oligonucleotides used in treatment of, e.g. pulmonary  
 PT vasoconstriction.  
 PS Disclosure; Page 37; 120pp; English.  
 XX  
 CC The specification describes antisense oligonucleotides (AA532869-X55271)  
 CC directed against at least 2 mRNAs selected from target genes, coding and  
 CC non-coding regions of RNAs corresponding to target genes, gene initiation  
 CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'-  
 CC end and the juxta-section between coding and non-coding regions and all  
 CC segments of RNAs encoding proteins associated with one or more diseases,  
 CC conditions or mixtures. The antisense oligonucleotides may be derived  
 CC from sequences AA55272-74. These multiple target oligonucleotides  
 CC (specifically AA55180-271) can be used for the antisense treatment of  
 CC diseases and conditions. Typical diseases and conditions are those  
 CC associated with impaired respiration and inflammation, including lung  
 CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,  
 CC acute asthma, allergies, asthma, impeded respiration, respiratory  
 CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,  
 CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary  
 CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.  
 CC colon cancer, breast cancer, lung cancer, pancreatic cancer,  
 CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as  
 CC well as all types of cancers which may metastasize or have metastasized  
 CC to the lungs, including breast and prostate cancer  
 XX  
 SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

Query Match 9.9%; Score 51.6; DB 2; Length 114955;  
 Best Local Similarity 33.7%; Pred. No. 0.74;  
 Matches 168; Conservative 49; Mismatches 273; Indels 9; Gaps 1;  
 QY 26 GCCCG 85  
 DB 104171 GCSNNNDNNGTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 104230  
 QY 86 GTTGACGACCGAGCTCATCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 145  
 DB 104231 DNNCCG 104290  
 QY 146 GTGACGACGCGGATGCGCGCGCTCGATC-----TCGACGCTCGCGCGCGCG 196  
 DB 104291 CG 104350  
 QY 197 CTCGTGCGCGCGCTCCACACCGAGGGGAACTGGCTCGTGGCAGCGGCGCGCGCG 256  
 DB 104351 CBGGCG 104410



PR 13-OCT-2000; 2000US-0239224P.  
 PR 12-APR-2001; 2001US-0283296F.  
 PR 24-JUL-2001; 2001US-00910813.  
 XX (ECOP-) ECOPIA BIOSCIENCES INC.  
 PA Farnet CM, Zazopoulos E, Staffa A;  
 XX  
 DR WPI; 2002-435445/46.  
 DR P-PSDB; AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,  
 DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,  
 DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,  
 DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
 DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.  
 XX  
 PT Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
 PT chemically modifying biological molecule that is a substrate for a  
 PT polypeptide encoded by a ramoplanin biosynthesis gene cluster.  
 XX  
 XX Disclosure; Page 87-135; 212pp; English.  
 XX  
 CC The invention relates to an isolated ramoplanin biosynthetic pathway  
 CC polypeptide selected from a polypeptide of open reading frames (ORF) 1-  
 CC 32. The isolated polypeptides are useful for chemically modifying a  
 CC biological molecule that is a substrate for a polypeptide encoded by a  
 CC ramoplanin biosynthesis gene cluster, by contacting the biological  
 CC molecule with the isolated polypeptide, where the polypeptide chemically  
 CC modifies the biological molecule. The method comprises contacting the  
 CC biological molecule with at least two different polypeptides encoded by  
 CC ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
 CC biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
 CC gene cluster comprising the ORFs is useful as a substrate for  
 CC bioengineering of antibiotic structures. An isolated polypeptide or its  
 CC encoding nucleic acid sequence is useful for generating derivatives of  
 CC ramoplanin, for improving production or for producing variants of other  
 CC antibiotics of the peptide class. The isolated polypeptides are useful  
 CC for synthesis of ramoplanin *in vivo* or *in vitro*, as an adenylation domain  
 CC in conjunction with other peptide synthetase modules and allowing the  
 CC incorporation of Thr into a peptide antibiotic precursor, for modifying  
 CC fatty acid structure and/or enhancing fatty acid incorporation into the  
 CC peptide antibiotic structure, for production of an hydroxyphenylglycine  
 CC (HPG)-containing peptide antibiotic, for enhancing secretion of  
 CC ramoplanin or its variants and derivatives, for enhancing uptake of  
 CC precursors for ramoplanin biosynthesis, for enhancing production of  
 CC ramoplanin products or its variants or derivatives, to chlorinate HPG of  
 CC a peptide antibiotic precursor, and for designing specific nucleotide  
 CC probes and primers for identifying and isolating putative lipopeptide  
 CC -producing microorganisms. This polynucleotide sequence represents the  
 CC 88421nt genomic DNA of a ramoplanin producing *Actinoplanes* sp.  
 CC microorganism of the invention  
 XX  
 SQ Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;  
 Query Match 9.8%; Score 51.2; DB 6; Length 88421;  
 Best Local Similarity 46.3%; Pred No. 0.89;  
 Matches 239; Conservative 0; Mismatches 273; Indels 4; Gaps 2;  
 7 GCACGACGACGCGGTGTCGCCCGCGCGCGGCGATGAGCGCGGTGCGCCCGCGAGCGCGA 67577  
 67518 GCACGACGACGTCGTGGCGCGCGCGCGCGGCGATGAGCGCGGTGCGCCCGCGAGCGCGA  
 67  
 67 CCAGGTGACGAGGTGGAGTTGACGACGAGCTCATCGCCGCTGCGCGCGATGCCGA 126  
 67578 ACTGCCCGCGCGGTGCGCGCGGAGCGCGGTGCGCGCGGTGCGCTGCGCTAGCGCGA 67637  
 127 AACACCGGCGCAGGATCAGTCAGCAGCGGATCGCGCTCGATCTCGACGCTCGGCC 186  
 67638 CGGCTCGGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 67697  
 187 GAGCTCGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 246  
 67698 CGCCCTGCG 67757

QY 247 GCCCAGCCGCGGACAGCTGCTCCAGGCTGCGCTCAACGACCTCTCCGCGGTCCGCGAGCGG 306  
 DB 67758 GCCGCTCTGCGCGCGCGCGCTGCGCGCTGCGCGAGCGCGCGCGCGCGCGCGCGCG 67817  
 QY 307 AAAGCGGACGCGCTCAGATCCCTGTCAGTCGCTCAGTCGCTCAGTCGCTCAGTCGCT 366  
 DB 67818 CGAGCTGCGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 67875  
 QY 367 GCCTGGGAGGATAGCGGTTTCAACGACGAGCGGACACACGCGCGCGCGCGCGCGCGCG 424  
 DB 67876 GCCAGGTCTCTCCGCCGATACGAGCGCGCGCGCGCTGCGGGCTCGCGCGCGCGCG 67935  
 QY 425 TTCAACGATTCCTCTCAGTACACGAGCGCTGCGCGGTGCGCGGTGCGCGGTGCGCGGT 484  
 DB 67936 TTGGCGCGCGCGCGCGCTGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCG 67995  
 QY 485 CCGCCCGGACCGCTCG 520  
 DB 67996 GCCCGCGCGGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 68031  
 RESULT 10  
 ABZ66676  
 ID ABZ66676 standard; DNA; 975 BP.  
 AC ABZ66676;  
 XX  
 DT 21-MAR-2003 (first entry)  
 XX  
 DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 14.  
 XX  
 KW Orthosomycin; biosynthesis; evernimicin; avilamycin; gene; ds.  
 XX  
 OS Streptomyces mobaraensis.  
 XX  
 FN WO200279505-A2.  
 PN 10-OCT-2002.  
 PD 28-MAR-2002; 2002WO-CA000432.  
 PF 28-MAR-2001; 2001US-0279095P.  
 PR 30-MAR-2001; 2001US-027909P.  
 PR 20-APR-2001; 2001US-0285214P.  
 XX  
 PA (ECOP-) ECOPIA BIOSCIENCES INC.  
 XX  
 PI Farnet CM, Zazopoulos E, Staffa A;  
 XX  
 DR WPI; 2003-058435/05.  
 DR P-PSDB; ABP99213.  
 XX  
 PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
 PT cluster, by detecting presence of nucleic acid sequence corresponding to  
 PT 17 of flammamycins protein families.  
 XX  
 PS Example 2; Page 173-174; 51pp; English.  
 XX  
 CC The invention relates to identifying orthosomycin biosynthetic genes and  
 CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
 CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
 CC ABP99362). The method is useful for identifying an orthosomycin  
 CC biosynthetic gene, gene fragment or gene cluster, especially an  
 CC evernimicin-type or avilamycin-type orthosomycin biosynthetic gene,  
 CC gene fragment or gene cluster. The method is useful for detecting the  
 CC presence of any organism that contains DNA for the production of  
 CC orthosomycins (both evernimicin-type orthosomycins and avilamycin-type  
 CC orthosomycins) regardless of the level at which genes for orthosomycin  
 CC production are expressed by the organism or the amount of orthosomycin  
 CC produced by the organism. This allows for the detection of new  
 CC orthosomycin natural products, not produced by the organism  
 XX  
 SQ Sequence 975 BP; 108 A; 378 C; 378 G; 111 T; 0 U; 0 Other;







Query Match 9.7%; Score 50.4; DB 7; Length 1209;  
Best Local Similarity 48.8%; Pred. No. 1.8; Mismatches 201; Indels 4; Gaps 2;  
Matches 195; Conservative 0;

QY 111 TCGCGCGCATCGGAACACCGGCGCAGGATCAGTCGACGACGCGCGATCGCGCTCG 170  
DB 989 TCGCGGTGTGCGGAACACGCGGCGGCGGACCATGGGACCCCGCGCGACCGCTCG 930  
QY 171 ATCTCGACGCTGGCGCGAGTCGATCTGTGTCGCGGCTCCACACAGGGGAACCTGG 230  
DB 929 CGCAGCTGTTGGCCCGCGCGGTGGGTGACGAACAGCTGGCGCATCCAGACGACGCGC 870  
QY 231 CTCGCTGGCAGGCGCAGCCCCCAGCGCGGACAGCTGTCAGGCTGCGCTCAGGACCTCT 290  
DB 869 TGGGGCGCCGCTCGACACGCGCAGCTGCGCGCGCGCGCGGAGC---CGGTGAC 813  
QY 291 CGCGGTTCGGGACGGAACCGCGCGCTCAGTCCCTGTGTCAGTCGATCGGCTCAGTG 350  
DB 812 GGGATCCCCCGGTGACACACCGCGCGGCGAG-TCCACTCGGACAGGCGCGCGCAG 754  
QY 351 CGGTGTCCTCCCTTGGCCTGGAGGATAGCGGTTCACGACGAGCGGACACGCGCGGC 410  
DB 753 TCGGCCAGGCTCTCACCGGTTCGGCGACGAGCGGTTCAGTCTGCGCGCGCACAG 694  
QY 411 GGGGGGGGGGTTTACCGCATTCGCTCGATGACGAGCGCTGCGGGGTGCGGGCGGT 470  
DB 693 GCGCGCCAGCGGATCCGACCGCGCCCGTACACACGCGGCGGAGCGGTCCAGTCCGC 634  
QY 471 GCGCGAGATCCGTACCGCGCGGACCGCTTCGCGCGCGCC 510  
DB 633 CACCCACGGGCGAGCGACCGCGCTGCGGCGACGCGCGC 594

RESULT 13  
ID ABZ77318/c  
ID ABZ77318 standard; cDNA; 2684 BP.  
XX AC ABZ77318;  
XX DT 28-MAY-2003 (first entry)  
XX DE Nucleotide sequence of a murine ACDK3 polypeptide.  
XX KW Human; ACDK; activator of cyclin-dependent kinase; cell division;  
XX KW Chromosome 10; ACDK3; cyclin dependent kinase; expressed sequence tag;  
XX KW EST; A1606540; gene; ss.  
XX OS Mus musculus.  
XX PH Key Location/Qualifiers  
XX FT CDS 72..2213  
XX FT /\*tag= a  
XX FT /product= "ACDK3"  
XX FT /transl\_except= (pos:2159..2144,aa:Thr)  
XX WO2003008557-A2.  
XX PN 30-JAN-2003.  
XX PD 19-JUL-2002; 2002WO-US023147.  
XX PF 19-JUL-2001; 2001US-0306835P.  
XX PA (UYFL ) UNIV FLORIDA.  
XX PI She J, Wang C, Kumar PG;  
XX DR WPI; 2003-239327/23.  
XX DR P-PSDB; ABP97757.  
XX PT New proteins and genes that are activators of cyclin-dependent kinase  
XX PT (ACDK), useful for regulating cell division, or for screening ACDK  
XX PT modulators for regulating the progression of a cell through the cell

PT cycle.  
XX Claim 25; Page 71-72; 103pp; English.  
XX CC The present sequence encodes a murine ACDK (activator of cyclin-dependent  
CC kinases) 3 polypeptide. This sequence is corresponds to expressed  
CC sequence tag (EST) A1606540. Human ACDK1 is localised to chromosome 10q23  
CC -10q24. ACDK polypeptides are predominantly localised within the nuclei  
CC of cells able to undergo mitosis. In cells that do not undergo mitosis,  
CC ACDK proteins are found in the cytoplasm. The ACDK protein or  
CC polynucleotide is useful for regulating cell division, and for regulating  
CC cyclin dependent kinase. They are also useful for screening modulators of  
CC ACDK expression or activity. These are useful for modulating ACDK  
CC expression or activity in cells to regulate the progression of a cell  
CC through the cell cycle

QY Sequence 2684 BP; 471 A; 834 C; 906 G; 473 T; 0 U; 0 Other;  
SQ Query Match 9.7%; Score 50.4; DB 7; Length 2684;  
Best Local Similarity 49.6%; Pred. No. 1.7;  
Matches 129; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 4 GCACGACGACGACCGGTGTGCGCGCGCGCCATCAGAACTGCACTCGCGCGCGG 63  
DB 780 GCCCGCGCAACGACGCGCGGCGACGCGGCTGGCCCGCGCCCGCTACAGCAGCA 721  
QY 64 TGACGAGTTCGACGAGTGGAGTTGAGAGCCAGCTCATCGCGCTGCGCGCGCATGC 123  
DB 720 CGGCGAGTGGCGCTGCGCGCGAGGTGGCCAGAAAGCAGCGCGCGCGCGCGG 661  
QY 124 CGAAACACCGCGCGCGGATCAGTGCAGCAGCGGATGCGCGCTCGATCTCGACGCTCG 183  
DB 660 CCCAGCGCGCGCGGTTCCAGCGCGCGCGCGCGCGCTCGCGCTCGAGCGCTCT 601  
QY 184 GCGCGAGTTCGATCTCGTCCCGCGCTCCACACGAGGAGAACTGGCTCGGTGGCAGCG 243  
DB 600 CGCGCAGCATCTGCACCTCGCGCGCGCGCGCGCGCGCTCAGTGCAGCGCGCGCG 541  
QY 244 GCAGCGCGCGCGGACAGC 263  
DB 540 CCACCGCGCGCGGCGCGC 521

RESULT 14  
ID AAL61224 standard; DNA; 82746 BP.  
XX AC AAL61224;  
XX DT 22-SEP-2003 (first entry)  
XX DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster I.  
XX KW Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX OS Actinosynnema pretiosum.  
XX WO2003045312-A2.  
XX PN 05-JUN-2003.  
XX PD 21-NOV-2002; 2002WO-US037547.  
XX PF 21-NOV-2001; 2001US-0332158P.  
XX PR (UNIW ) UNIV WASHINGTON.  
XX PA Floss HG, Yu T, Leistner E;  
XX PI WPI; 2003-493374/46.  
XX PT Novel maytansinoid produced by bacterial host cell transformed with  
XX PT expression vector comprising open reading frame from ansamitocin gene



```
QY 297 TCGGAGCGGAAACCGGACGGCTCAGATCCTGTCTAGTCGGCTCAGTCCGGTTC 356
Db 2392 CAGAGGTTTCAGCGCGGCAACCTCTGCTCCGCTGGTGCACAGGGTGACCGGACCTG 2451
QY 357 GTCCTCCCTTGGCTCGGAGGATAGCGGTTTCACGACGAGCGGCACCA 402
Db 2452 GCGGCACACGACGCGCGCGGAGGGTCTACACACGCGGCACCA 2497

RESULT 16
AAL61221/c
ID AAL61221 standard; DNA; 1554 BP.
AC AAL61221;
XX
DT 22-SEP-2003 (first entry)
DE Actinosynnema pretiosum transcriptional activator gene #2.
KW Maytansinoid; ansamitocin; antitumour; transcriptional activator; gene;
DS.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
XX WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Disclosure; Page 103-104; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumor agent. The present sequence is A. pretiosum
CC ansamitocin gene cluster II transcriptional activator gene
XX
SQ Sequence 1554 BP; 157 A; 520 C; 704 G; 173 T; 0 U; 0 Other;

Query Match 9.5%; Score 49.6; DB 7; Length 1554;
Best Local Similarity 47.3%; Pred. No. 2.5;
Matches 229; Conservative 0; Mismatches 239; Indels 16; Gaps 2;

QY 37 CCATCAGCAACTGCCACTCCGCGGGGTACACAGTCCAGTCCAGTGGAGTTGACGCC 96
Db 553 CCGACAGCAGCTGCTCGCACCGAGGCGGCCACGCTCCGCGGGTCCGAGTACAGCAGC 494
QY 97 AGCTCATCCCGCTCGCGCGGCATGCCGAAACACCGGCCAGGATCAGTGCAGCAGCG 156
Db 493 CGGACAGCGCAGGACGCGCGCACCGCGAGCGCGCGGGTCCCGCGCCAGCCGCACT 434
QY 157 CCATCGCGCCCTCGATCTCGAGTGTGCGCGCGGAGCTCGATCTCGTCTCGCGCGCTCCACA 216
Db 433 CCGCGAGCACCGCCTCCGCGCGGTCCGCGCACGCTGTGCGCGTCCCGCGCCACCAAGCAGT 374
QY 217 CCAGGGGAAACTGGCTCGGTGGCAGCGGAGCGCCAGCGCGGACAGCTGTCAGGCTGG 276
Db 373 CCAGGAGCACCGGTTTCGGCCCTCGGCGCTGTCTCGCCCTGTGTGCGCGC-----GG 323
QY 277 CGTCAAGCACTTCTCGCGGGTCCGGAGCGGAAACGCGCACCGCTCAGATCCCTGTGATC 336
```

```
Db 322 AGTCCACGTCCTCCCGCGCGCGGT-----CCCCGGCTTGGCCGACACCGGGGTAGCG 270
QY 337 GCATCGGCTCAGTGGCGGTCTGCTCCCTTGGCTTGGAGGATAGCGGTTTCACGACGAGCG 396
Db 269 CCAGCCAGCGAGTCTCCAGCACCGGACCTCTCTCGCGCGCGCGGTCTGGCCCGCAGC 210
QY 397 GCACCAAGCGCGCGCGGGGGGTTTACGCCGATCCGCTCGATGACCAAGCGGCTGC 456
Db 209 CCGCGCAGCACCGCGTCCGAGTCCGCGGTCTCCCGGCCACGCGCAGGTGCTCCACGAGC 150
QY 457 GGGGTCCGGCGGTTCGCGCAGATCCGTACCGCCCGACCGCTCCGGCCAGCGCGCGCGG 516
Db 149 GTCGCGCTGTGCGCACCGGACACCTGCTCCCGCGCGGAAACGCGCCAGCGCGGTGTGG 90
QY 517 ATCC 520
Db 89 TGCC 86

RESULT 17
AAL61225
ID AAL61225 standard; DNA; 11905 BP.
XX
AC AAL61225;
XX
DT 22-SEP-2003 (first entry)
DE Actinosynnema pretiosum ansamitocin biosynthetic gene cluster II.
XX
KW Maytansinoid; ansamitocin; antitumour; gene; ds.
XX
OS Actinosynnema pretiosum.
XX
PN WO2003045312-A2.
XX
PD 05-JUN-2003.
XX
PF 21-NOV-2002; 2002WO-US037547.
XX
PR 21-NOV-2001; 2001US-0332158P.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Floss HG, Yu T, Leistner E;
XX
XX WPI; 2003-493374/46.
XX
PT Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
PS Disclosure; Page 153-160; 160pp; English.
XX
CC The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
CC useful as a potent antitumor agent. The present sequence is A. pretiosum
CC ansamitocin ansamitocin biosynthetic gene cluster II
XX
SQ Sequence 11905 BP; 1346 A; 4416 C; 4697 G; 1446 T; 0 U; 0 Other;

Query Match 9.5%; Score 49.6; DB 7; Length 11905;
Best Local Similarity 47.3%; Pred. No. 2.1;
Matches 229; Conservative 0; Mismatches 239; Indels 16; Gaps 2;

QY 37 CCATCAGCAACTGCCACTCCGCGGGGTGACAGGTCCAGTCCAGTGGAGTTGACGAGCC 96
Db 11352 CCGACAGCAGCTGCTCGCACCGAGGAGCGCGGCCAGCTCCGCGGTCCGAGTACAGCAGCG 11411
QY 97 AGCTCATCCCGCTCGCGCGGCATGCCGAAACACCGGCCAGGATCAGTGCAGCAGCG 156
Db 11412 CGGACAGCGCAGGACGCGCGGCCACCGCGCGCGGGTCCCGCGCCAGCGCGCACT 11471
```









FT	partial	/product= "EvrQ"	complement (38072. .38566)	FT	CDS
FT	/note= "No start codon"	/tag= at		FT	
FT	complement (12108. .13022)	/product= "EvrR"		FT	CDS
FT	/tag= s	complement (38892. .40163)		FT	
FT	product= "EvdL"	/tag= au		FT	CDS
FT	complement (13027. .13030)	/product= "EvrS"		FT	
FT	/tag= t	complement (40216. .40890)		FT	
FT	product= "EvdT"	/tag= av		FT	CDS
FT	/tag= u	/product= "EvrT"		FT	
FT	complement (15369. .15373)	complement (40887. .41576)		FT	CDS
FT	/tag= v	/tag= ax		FT	
FT	complement (15380. .16414)	/product= "EvrU"		FT	RBS
FT	/tag= w	complement (40899. .40902)		FT	
FT	product= "EvrB"	/tag= aw		FT	CDS
FT	complement (16419. .17873)	/tag= ay		FT	
FT	/tag= x	/product= "EvrV"		FT	RBS
FT	product= "EvrC"	complement (42714. .42717)		FT	
FT	complement (17870. .18934)	/tag= az		FT	CDS
FT	/tag= y	complement (42810. .43799)		FT	
FT	product= "EvrD"	/tag= ba		FT	CDS
FT	19374. .20906	/product= "EvrW"		FT	
FT	/tag= z	complement (43799. .44866)		FT	CDS
FT	product= "EvrE"	/tag= bc		FT	
FT	21056. .22542	/product= "EvrX"		FT	RBS
FT	/tag= ab	complement (43807. .43811)		FT	
FT	21064. .22542	/tag= bb		FT	CDS
FT	/tag= aa	complement (45014. .45760)		FT	
FT	product= "EvrF"	/tag= bd		FT	RBS
FT	22736. .22740	/product= "EvrY"		FT	
FT	/tag= ad	complement (45767. .45770)		FT	RBS
FT	22748. .24172	/tag= be		FT	
FT	/tag= ac	complement (45952. .45956)		FT	RBS
FT	product= "EvrG"	/tag= bg		FT	CDS
FT	complement (24177. .25223)	complement (45962. .46714)		FT	
FT	/tag= ae	/tag= bf		FT	
FT	product= "EvrH"	/product= "EvrZ"		FT	CDS
FT	complement (25230. .25233)	complement (47156. .49234)		FT	
FT	/tag= af	/tag= bh		FT	
FT	25550. .26626	/product= "EvsA"		FT	CDS
FT	/tag= ag	51627. .52715		FT	
FT	product= "EvrI"	/tag= bi		FT	
FT	26672. .26676	/product= "EvsB"		FT	RBS
FT	/tag= ai	51629. .51622		FT	
FT	26685. .30479	/tag= bj		FT	CDS
FT	/tag= ah	52889. .53857		FT	
FT	product= "EvrJ"	/tag= bk		FT	CDS
FT	complement (30557. .31876)	/product= "EvsC"		FT	
FT	/tag= aj	53554. .54207		FT	
FT	product= "EvrK"	/tag= bl		FT	CDS
FT	complement (31885. .31888)	/product= "EvsA"		FT	
FT	/tag= ak	complement (54362. .55117)		FT	CDS
FT	complement (31941. .32882)	/tag= bm		FT	
FT	/tag= al	/product= "EvsB"		FT	RBS
FT	product= "EvrL"	complement (55125. .55128)		FT	
FT	complement (33167. .34405)	/tag= bn		FT	CDS
FT	/tag= am	complement (55135. .56094)		FT	
FT	product= "EvrM"	/tag= bo		FT	
FT	complement (34414. .34418)	/product= "EvsC"		FT	RBS
FT	/tag= an	complement (56100. .56103)		FT	
FT	complement (34449. .35210)	/tag= bp		FT	CDS
FT	/tag= ao	complement (56184. .56613)		FT	
FT	product= "EvrN"	/tag= bq		FT	
FT	complement (35219. .35221)	/product= "EvsC2"		FT	
FT	/tag= ap			FT	
FT	complement (35294. .36238)			FT	
FT	/tag= aq			FT	
FT	product= "EvrO"			FT	CDS
FT	complement (36235. .36963)			FT	
FT	/tag= ar			FT	
FT	product= "EvrP"			FT	CDS
FT	complement (36998. .38026)			FT	
FT	/tag= as			FT	

Query Match 9.4%; Score 49; DB 5; Length 109519;  
Best Local Similarity 51.1%; Pred. No. 2.2;  
Matches 115; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 29 CCGCGCGCCATCAGCACTCCGCGGGGTGACCGTCCAGTCCAGTGGAGTT 88  
DB 82792 CCGCTCAGCTAGACGTCGCGGGTTCGTCGGGTCCGCGCGGAGCGGAGTTC 82733



```
QY      89 GAGCAGCCAGCTCATCGCGCGCTGGCGCGCATGCCGAAACACACCGGCGCAGGATCACGTG 148
      |||||
Db      82732 GAGCAGCAGACAGTGCCTCGGCTCGCGCGCAGCGCGCGCGCGCGCAGCGGAC 82673
      |||||
QY     149 CAGCAGCGGATCGCGCGCTCGATCTCGACGCTCGCGCGCAGCTCGATCTCGTCCCGGG 208
      |||||
Db     82672 CGGTACGCTGATGTTGTGCGCTTCGCGCGCAGCGCGCAGGTCGCGAGCTGTCGGCCAT 82613
      |||||
QY     209 CTCGCCACACGAGGGAACCTGCTCGGTGGCGAGCGCGCAGCCCCAG 253
      |||||
Db     82612 CACCGATCGCGCGCGCAGCGCGCGCGCGCAGCGCGCTTCGCTGAG 82568
      |||||

RESULT 24
ID      AAD54645/c
XX      AAD54645 standard; DNA; 113193 BP.
AC      AAD54645;
XX      26-JUN-2003 (first entry)
XX      Streptomyces nodosus amphotericin (amph) biosynthetic gene cluster.
DE      Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
KW      Streptomyces nodosus.
XX      Location/Qualifiers
FH      complement(4. .1824)
FT      /tag= a
FT      /product= "ABC transporter encoded by S. nodosus amphG
FT      gene"
FT      complement(1805. .3628)
FT      /tag= b
FT      /product= "ABC transporter encoded by S. nodosus amphH
FT      gene"
FT      3840. .4874
FT      /tag= c
FT      /product= "GDP-mannose dehydratase encoded by S. nodosus
FT      amphDII gene"
FT      5042. .33574
FT      /tag= d
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
FT      nodosus amphI gene"
FT      33584. .50518
FT      /tag= e
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 15, 16 and 17 encoded by S. nodosus
FT      amphJ gene"
FT      50571. .56675
FT      /tag= f
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 18 and thioesterase encoded by S.
FT      nodosus amphK gene"
FT      56829. .58019
FT      /tag= g
FT      /product= "Cytochrome P450 encoded by S. nodosus amphL
FT      gene"
FT      58139. .58648
FT      /tag= h
FT      /product= "ORF1, hypothetical protein"
FT      complement(58756. .59610)
FT      /tag= i
FT      /product= "ORF2, hypothetical protein"
FT      59869. .61470
FT      /tag= j
FT      /product= "ORF3, hypothetical protein"
FT      /transl_except= (pos:59869. .59871, aa:Met)
FT      complement(61798. .61995)
FT      /tag= k
FT      /product= "Ferredoxin encoded by S. nodosus amphM gene"
```

```
FT      CDS      complement(62051. .63250)
FT      /tag= l
FT      /product= "Cytochrome P450 encoded by S. nodosus amphN
FT      gene"
FT      complement(63250. .64308)
FT      /tag= m
FT      /product= "NDP-sugar aminotransferase encoded by S.
FT      nodosus amphDII gene"
FT      complement(64324. .65775)
FT      /tag= n
FT      /product= "Glycosyl transferase encoded by S. nodosus
FT      amphDI gene"
FT      /transl_except= (pos:65773. .65775, aa:Met)
FT      60801. .70319
FT      /tag= o
FT      /product= "Polyketide synthase multienzyme housing
FT      loading module encoded by S. nodosus amphA gene"
FT      70366. .79938
FT      /tag= p
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 1 and 2 encoded by S. nodosus by amphB
FT      gene"
FT      79956. .112709
FT      /tag= q
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT      nodosus by amphC gene"
XX      WO200297082-A2.
XX      05-DEC-2002.
XX      27-MAY-2002; 2002WO-IB0000071.
XX      31-MAY-2001; 2001IE-00000527.
XX      (UYDU-) UNIV COLLEGE DUBLIN.
XX      Caffrey JP;
XX      WFI; 2003-201271/19.
XX      P-PSDS; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX      AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
XX      AAE36129, AAE36130, AAE36131, AAE36132.
XX      Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX      for preparing amphotericin derivative or analog antibiotic agent with
XX      altered properties, in biosynthesis of polyketide other than
XX      amphotericin.
XX      Claim 1; Page 52-114; 276pp; English.
XX      The invention relates to the gene cluster encoding the polypeptides
XX      responsible for the biosynthesis of the polyene antibiotic amphotericin
XX      (amph) of Streptomyces nodosus. Polynucleotides of the invention are
XX      useful for preparing amphotericin derivatives or analogue antibiotic
XX      agents with altered properties and in the biosynthesis of polyketides
XX      other than amphotericin. amphDIII, amphDII or amphDI mutants are useful
XX      for producing amphotericin derivatives glycosylated with alternative
XX      sugars; amphDIII or amphDII gene sequences are useful in engineered
XX      biosynthesis of perosaminyl-amphoteronolide B; amphDIII or amphDII and
XX      amphN gene sequences are useful in the engineered biosynthesis of
XX      perosaminyl-16-descarboxyl-16-methyl amphoteronolide B; amphDIII, amphDII
XX      and amphDI gene sequences are useful for preparing polypeptides capable
XX      of addition of mycosamine to a polyketide other than amphoteronolide A or
XX      B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX      The present sequence is S. nodosus amph biosynthetic gene cluster
XX      Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
XX      Query Match 9.4%; Score 49; DB 7; Length 113193;
XX      Best Local Similarity 48.1%; Pred. NO. 2.2;
XX      Matches 139; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
```



PI Floss HG, Yu T, Leistner E;  
XX WPI; 2003-493374/46.  
XX Novel maytansinoid produced by bacterial host cell transformed with  
PT expression vector comprising open reading frame from ansamitocin gene  
PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
XX Disclosure; Page 102; 160pp; English.  
XX  
XX The invention relates to maytansinoid produced by bacterial host cell  
CC transformed with expression vector comprising open reading frame from  
CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
CC useful as a potent antitumor agent. The present sequence is A. pretiosum  
CC ansamitocin gene cluster II Am46 gene  
XX  
XX Sequence 795 BP; 63 A; 306 C; 338 G; 88 T; 0 U; 0 Other;  
SQ  
Query Match 9.3%; Score 48.6; DB 7; Length 795;  
Best Local Similarity 47.3%; Pred. No. 3.9;  
Matches 219; Conservative 0; Mismatches 234; Indels 10; Gaps 2;  
QY 49 GCACCTCCGGGGGTGACAGTCCAGCGTGGAGTTGAGCAGCGCTCATCGCG 108  
Db 661 GCGGTTCGCGCGCGGTGCGAGCTCGCGCGATCATCTGTCCAGTAGCGCGCCACG 602  
QY 109 CCGTGGCGCGCATGCCGAAACACACCGCGGCGCAGGATCACGTGACGACGCGGATGCGCGCT 168  
Db 601 CGCGCGCGAGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 542  
QY 169 CGATTCGACGCTCGCGCGAGCTCGATCTGTGCGCGCGCTCCACACCGAGGGAAC 228  
Db 541 CGGTGCGCGCGCGCGCGGTGAGCGGTCTGCCCGAGCTTCGCGAGTGGAAACGGAAC 482  
QY 229 GCGTGGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 288  
Db 481 CGGGCTGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 428  
QY 289 CTCGCGGTGCGGAGCGGAAACCGCGACGCTCAGATCCCTGTAGTCGATCGGCTCAG 348  
Db 427 CGGTGCG 368  
QY 349 TCGCGTCTGCTCCCTTGGCTGGGAGGATAGCGGTTCACGACGAGCGGCGACCGCGG 408  
Db 367 CGACGTCAGCGCGCGCGGTTCGCTGGAGCGCGGTGCGACGCGCTCGACCGGAC 308  
QY 409 GCGGGGCGCGCGCGGTTCAGCGGATCGCGCTCGATGACGCGGTGCGGGGTGCGGGCG 468  
Db 307 CGCGCGGCAAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 252  
QY 469 GTGCGGAGATCGGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511  
Db 251 GCGCGAGCAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 209

RESULT 27  
ACA40900/C  
ID ACA40900 standard; DNA; 1797 BP.  
XX ACA40900;  
XX  
XX 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #22557.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX Mycobacterium tuberculosis.  
XX WO200277183-A2.  
XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR 05-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU37030.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 28770; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway;  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than S. aureus, S. typhimurium,  
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1797 BP; 175 A; 807 C; 616 G; 199 T; 0 U; 0 Other;  
SQ  
Query Match 9.3%; Score 48.6; DB 7; Length 1797;  
Best Local Similarity 47.7%; Pred. No. 3.7;  
Matches 175; Conservative 0; Mismatches 189; Indels 1; Gaps 1;  
QY 154 CGCGGATGCGCGCTCGATCTCGACGCTGCGCGCGAGCTCGATCTCGTCCCGGCTCC 213  
Db 1768 CGCGGCTGCTCGCGCGCGATCAACGAGTCTTCTCGGCGAATACCGGCGCGCTGATCG 1709  
QY 214 ACACCGGGGAACTGCTCGGTGGGAGCGGCGGACCGCGCGCGGACAGCTCTCCAGGC 273  
Db 1708 GCAACGCGACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTAA 1649  
QY 274 TGGCGTCAACGACCTCTCGCGGTCGCGGAGCGGAGAAACGCGACGCTCAGATCCCTGTCA 333  
Db 1648 TCGGCAACGCGCGCGCGCGGAGGATCCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1589

QY 334 GTGCTATCGGCTCAGTGGGCTGCTCCCTTGGCTGGAGGATAGCGTTTACAGCA 393  
Db 1588 ACAGCGGCGCGGCTCATCGGCAACGGCGGCGCGCGCGCGCGGAGGCGCA 1529  
QY 394 GCGGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 453  
Db 1528 GCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1470  
QY 454 TGCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 513  
Db 1469 GCGGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1410  
QY 514 GGG 516  
Db 1409 GGG 1407

RESULT 28  
AAL61225/c  
ID AAL61225 standard; DNA; 11905 BP.  
XX AC  
AC AAL61225;  
XX

22-SEP-2003 (first entry)  
DT  
XX Actinosynnema pretiosum ansamitocin biosynthetic gene cluster II.  
DE  
XX Maytansinoid; ansamitocin; antitumour; gene; ds.  
XX  
XX Actinosynnema pretiosum.  
XX  
XX WO2003045312-A2.  
XX  
XX 05-JUN-2003.  
XX  
XX 21-NOV-2002; 2002WO-US037547.  
XX  
XX 21-NOV-2001; 2001US-0332158P.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Floss HG, Yu T, Leistner B;  
XX  
XX WPI; 2003-493374/46.

Novel maytansinoid produced by bacterial host cell transformed with  
expression vector comprising open reading frame from ansamitocin gene  
cluster I of Actinosynnema pretiosum, useful as antitumor agent.  
XX  
XX Disclosure; Page 153-160; 160pp; English.  
XX  
XX The invention relates to maytansinoid produced by bacterial host cell  
transformed with expression vector comprising open reading frame from  
ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is  
useful as a potent antitumor agent. The present sequence is A. pretiosum  
XX  
XX ansamitocin ansamitocin biosynthetic gene cluster II  
XX  
SQ Sequence 11905 BP; 1346 A; 4416 C; 4697 G; 1446 T; 0 U; 0 Other;

Query Match 9.38; Score 48.6; DB 7; Length 11905;  
Best Local Similarity 47.38; Pred. No. 3.1;  
Matches 219; Conservative 0; Mismatches 234; Indels 10; Gaps 2;  
QY 49 GCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 108  
Db 8884 GCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 8825  
QY 109 CCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCT 168  
Db 8824 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCT 8765  
QY 169 CGATCTCGAGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 228

Db 8764 CGGTGCGCGCGCGCGGCTGAGCGGTCTGCCCGCAGCTCTGCCAGCAGTGAACAGGAACG 8705  
QY 229 GGCTCGGTGGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 288  
Db 8704 CGGCGCTGCGCGGTTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8651  
QY 289 CTCGCGGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 348  
Db 8650 CGGTGCGCTGCG 8591  
QY 349 TGGCGGTGCTCCCTTGGCGCTGGAGGATAGCGGTTTCAGACAGCGGCGGCGGCGG 408  
Db 8590 CGAGCTCCAGCGCCAGTTCGCTGGAGCGCGTTCGCGACCTCGACAGCGGCTCGACCGGAC 8531  
QY 409 GCGGCGGCGGCGGCGGCGGTTTCAGCCGATCCGCTTCGATGACCGCGCTCGGCGGCGG 468  
Db 8530 GCGCGGCGCAAGCCTCTCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8475  
QY 469 GTCGCGGAGATCCGTACCGCGCGCGCGCGCGCGCTCGGCGGCGGCGGCGGCGG 511  
Db 8474 GCGCGGCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8432

RESULT 29  
AAL99683\_03/c  
Continuation (4 of 44) of AAL99683 from base 300001 (Mycobacterium tuberculosis strain H  
WP Sequence split into 44 fragments LOCUS AAL99683 Accession AAL99683  
WP Fragment Name Begin End  
WP AAL99683\_00 1 110000  
WP AAL99683\_01 100001 210000  
WP AAL99683\_02 200001 310000  
WP AAL99683\_03 300001 410000  
WP AAL99683\_04 400001 510000  
WP AAL99683\_05 500001 610000  
WP AAL99683\_06 600001 710000  
WP AAL99683\_07 700001 810000  
WP AAL99683\_08 800001 910000  
WP AAL99683\_09 900001 1010000  
WP AAL99683\_10 1000001 1110000  
WP AAL99683\_11 1100001 1210000  
WP AAL99683\_12 1200001 1310000  
WP AAL99683\_13 1300001 1410000  
WP AAL99683\_14 1400001 1510000  
WP AAL99683\_15 1500001 1610000  
WP AAL99683\_16 1600001 1710000  
WP AAL99683\_17 1700001 1810000  
WP AAL99683\_18 1800001 1910000  
WP AAL99683\_19 1900001 2010000  
WP AAL99683\_20 2000001 2110000  
WP AAL99683\_21 2100001 2210000  
WP AAL99683\_22 2200001 2310000  
WP AAL99683\_23 2300001 2410000  
WP AAL99683\_24 2400001 2510000  
WP AAL99683\_25 2500001 2610000  
WP AAL99683\_26 2600001 2710000  
WP AAL99683\_27 2700001 2810000  
WP AAL99683\_28 2800001 2910000  
WP AAL99683\_29 2900001 3010000  
WP AAL99683\_30 3000001 3110000  
WP AAL99683\_31 3100001 3210000  
WP AAL99683\_32 3200001 3310000  
WP AAL99683\_33 3300001 3410000  
WP AAL99683\_34 3400001 3510000  
WP AAL99683\_35 3500001 3610000  
WP AAL99683\_36 3600001 3710000  
WP AAL99683\_37 3700001 3810000  
WP AAL99683\_38 3800001 3910000  
WP AAL99683\_39 3900001 4010000  
WP AAL99683\_40 4000001 4110000  
WP AAL99683\_41 4100001 4210000  
WP AAL99683\_42 4200001 4310000  
WP AAL99683\_43 4300001 4403765

Query Match		9.3%; Score 48.6; DB 4; Length 110000;
Best Local Similarity		47.7%; Pred. No. 2.5; Mismatches 189; Indels 1; Gaps 1;
Matches 173; Conservative		0;
QY	154	CGGCGATCGCCCTCGATCTCGACGGTGGCGCGAGCTCGATCTCGTCCCGGCTCCC 213
Db	38841	CGCCGCTGCTCGCCCGGATCAACAGAGTTCTTCTGCGCAATACCGGGCGCCGCTGATCG 38782
QY	214	ACACAGGGGAACCTGGTCTGGTGCAGCGGAGCCAGCCAGCGGAGACAGCTCTCCAGGC 273
Db	38781	GCACCGCAACCAACGGCCCGCCCGGACACCGGGCCCAATGGCGGGAGCCGGCTGGTTAA 38722
QY	274	TGGCGTCAACGACCTCTCGCGGTCGGGAGCGGAAACGCGACGCTCAGATCCCTGTCA 333
Db	38721	TGGCAACGCGCGCGGAGGATCCGGCGCGCGCGGCGTCAACGGCGCGCGCGCGCA 38662
QY	334	GTGCGATCGGCTCAGTCCGCTCCCTTGGCTGGGAGGATAGCGGTTACAGCA 393
Db	38661	ACGGCGGCGCGCGCGGCTCTATCGGCAACGGCGGCGCGCGCGCGCGCGCGGAGGCA 38602
QY	394	GGGCAACCAACGCGCGCGCGGCGCGGCGGTTCAAGCGGATCCGCTCGATGACACGCGGC 453
Db	38601	GCACGGGACCG 38543
QY	454	TGCGGGTGGCGCGGTCGGCGAGATCCGTACCGCGCGGACCGCTCGGCGACGCGCGCC 513
Db	38542	GCCGGGGTGGCGGTCGGCGGATTCGAGCGCGCTTCGGCGCGCACCGCGCGCGCGCGC 38483
QY	514	GGG 516
Db	38482	GGG 38480
RESULT 30		
AAI99683_39		Continuation (40 of 44) of AAI99683 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 44 fragments		LOCUS AAI99683 Accession AAI99683
WP	Fragment Name	Begin End
WP	AAI99683_00	1 110000
WP	AAI99683_01	100001 210000
WP	AAI99683_02	200001 310000
WP	AAI99683_03	300001 410000
WP	AAI99683_04	400001 510000
WP	AAI99683_05	500001 610000
WP	AAI99683_06	600001 710000
WP	AAI99683_07	700001 810000
WP	AAI99683_08	800001 910000
WP	AAI99683_09	900001 1010000
WP	AAI99683_10	1000001 1110000
WP	AAI99683_11	1100001 1210000
WP	AAI99683_12	1200001 1310000
WP	AAI99683_13	1300001 1410000
WP	AAI99683_14	1400001 1510000
WP	AAI99683_15	1500001 1610000
WP	AAI99683_16	1600001 1710000
WP	AAI99683_17	1700001 1810000
WP	AAI99683_18	1800001 1910000
WP	AAI99683_19	1900001 2010000
WP	AAI99683_20	2000001 2110000
WP	AAI99683_21	2100001 2210000
WP	AAI99683_22	2200001 2310000
WP	AAI99683_23	2300001 2410000
WP	AAI99683_24	2400001 2510000
WP	AAI99683_25	2500001 2610000
WP	AAI99683_26	2600001 2710000
WP	AAI99683_27	2700001 2810000
WP	AAI99683_28	2800001 2910000
WP	AAI99683_29	2900001 3010000
WP	AAI99683_30	3000001 3110000
WP	AAI99683_31	3100001 3210000
WP	AAI99683_32	3200001 3310000
WP	AAI99683_33	3300001 3410000
WP	AAI99683_34	3400001 3510000

Query Match		9.3%; Score 48.6; DB 4; Length 110000;
Best Local Similarity		45.5%; Pred. No. 2.5; Mismatches 249; Indels 1; Gaps 1;
Matches 209; Conservative		0;
QY	57	GGCGGGGTGACAGGTCGACCGAGTTCAGCAGCGCAGCTCATCGCCGCTCGGCC 116
Db	25089	GGGGTGTATCAATCGGCCACCCAGGCGCTGCTGGGGCGCCGTTGATCGGTGACGGCGCC 25148
QY	117	GGCATGCGGAAAACACCGGGCCAGGATCATGTCAGCACGGCGATCGCGGCTCGATCTCG 176
Db	25149	AATGCCAGCACTCCCGCGGGCGCGCGGGCGCGGGTCTGCTGTTTCGGCAACGGCGGG 25208
QY	177	ACGTTGGCGCGAGCTCGATCTCGTCGCCCGGCTCCACACAGGGGAAACTGGCTCGGT 236
Db	25209	GCUGGGCAGCCGGGGCGCCCGCCAGCGCGGCTGGCGGGCGCCCGGATTTGG 25268
QY	237	GGCAGCGGACGCCCGACCGCGGACAGCTCGTCCAGCTGCGCTGTCACGGACCTCTCGCGG 296
Db	25269	GGCAACGGCGGCGCGCGGGCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25328
QY	297	TCGGGAGCGGAAACGCGCAGCGCTCAGATCCCTGTCAGTCGCTCAGTCGCTCAGTCCGCTC 356
Db	25329	GCCTGGTGGTGGTCTTTCGGGGTTCGGGGCGCGCGGGTGTCTAGTGGGGCGGTGGCGGC 25388
QY	357	GTCCCTTGGCTGGGAGGATAGCGGTTTCACGACGAGCGGCAACCGCGGCGCGGGCGG 416
Db	25389	ACCGCGGGCGCGCGCGCGCGCGCGGCTGTTGATCTGGGGCGCGCGCGCGCGCGGCTGTC 25448
QY	417	GGGGGGGTTTCAGCGGATCCGCTCGATGACCGAGCGCTGCGGGTGGGGCGGTGGCGCA 476
Db	25449	GGTGGGGCGGTGGCGGACCGCGGGG-CGCGCGCGCGCGCGGAGCTGCTGTTCGGCGC 25507
QY	477	GATCCGTACCGCGCGGACCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 515
Db	25508	CGGCGGTGCGGGTGGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 25546
RESULT 31		
ACA27331/c		
ID	ACA27331	standard; DNA; 1689 BP.
XX	ACA27331;	
XX	ACA27331;	
DT	19-JUN-2003	(first entry)
XX	Prokaryotic essential gene #8988.	
XX	Antisense; ds; prokaryotic essential gene; cell proliferation;	
KW	drug design; gene.	
XX	Bordetella pertussis.	
XX	WO200277183-A2.	
FN	03-OCT-2002.	
PD	03-OCT-2002.	
XX	21-MAR-2002; 2002WO-US009107.	
PF	21-MAR-2001; 2001US-00815242.	
XX	06-SEP-2001; 2001US-00948993.	
PR	25-OCT-2001; 2001US-0342923P.	
PR	08-FEB-2002; 2002US-00072851.	
PR	06-MAR-2002; 2002US-0362699P.	

XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-029926/02.  
XX DR P-PSDB; ABU23461.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 14; SEQ ID NO 15201; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX CC prokaryotic essential genes. Note: The sequence data for this patent did  
XX CC not form part of the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1689 BP; 221 A; 626 C; 556 G; 286 T; 0 U; 0 Other;  
Query Match 9.3%; Score 48.4; DB 7; Length 1689;  
Best Local Similarity 53.8%; Pred. No. 4;  
Matches 100; Conservative 0; Mismatches 86; Indels 0; Gaps 0;  
Qy 9 ACAGACGACCGGTGGTCCCGCCGCGCCGATCAGCACTGCCACTCGCGGGGTGACC 68  
Db 1391 AAGAACAGCGCCCTTCGACCGCGTAGACGACGTCGTCGCGCAGCGCCGTAGCCG 1332  
Qy 69 AGGTCGACAGGTGGGAGTTGAGCAGCGCAGCTCATCGCGCTCGCGCGCATGCCGAAA 128  
Db 1331 CGGTGGCCCACTGTCTGTCGCCAGCTCGCGCGGCTCTGCGGCTGCACTTCGACGCTGGAC 1272  
Qy 129 CACGGGCGCAGATCACTGACGACGCGGATCGCGCTCGATCTCGACGCTGGCGCG 198  
Db 1271 GCCATGCGCCGACGAGGATGACGATGCGGATGACGCGGATGTTGACCGCCACC 1212  
Qy 189 AGCTCG 194  
Db 1211 AGGTGCG 1206  
RESULT 32  
ACA38077/c  
ID ACA38077 standard; DNA; 3669 BP.

XX ACA38077;  
XX 19-JUN-2003 (first entry)  
XX DE Prokaryotic essential gene #19734.  
XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX KW drug design; gene.  
XX OS Mycobacterium avium.  
XX XX WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX XX WPI; 2003-029926/02.  
XX DR P-PSDB; ABU34207.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 14; SEQ ID NO 25947; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX CC prokaryotic essential genes. Note: The sequence data for this patent did  
XX CC not form part of the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 3669 BP; 540 A; 1243 C; 1388 G; 498 T; 0 U; 0 Other;  
Query Match 9.3%; Score 48.4; DB 7; Length 3669;

XX	DR	WPI; 2002-599794/64.
XX	PT	Isolated nucleic acid molecule from a bacterial daptomycin biosynthetic gene cluster encoding a thioesterase or thioesterase domain, useful for generating novel linear and cyclic peptides, and products in a cell.
PT	PT	Claim 7; Page 142-165; 227pp; English.
PT	PT	The invention relates to a novel isolated nucleic acid molecule comprising a sequence that encodes a thioesterase or thioesterase domain, derived from a bacterial daptomycin biosynthetic gene cluster. The proteins of the invention have antibacterial, fungicide, virucide, antiparasitic, immunomodulator, antilipemic, and cytostatic activity. The polynucleotides may have a use in gene therapy. The compositions and methods of the present invention are useful for generating novel linear and cyclic peptides and improving yield of a product in a cell expressing an daptomycin non-ribosomal peptide synthetase (NRPS) to be used as new compounds or in producing new compounds, such as antibiotics, antifungals, antivirals, antiparasitics, antimitotics, antitumour agents, immunomodulatory agents, anti-cholesterolemic agents, siderophores, agrochemicals and cytostatics. The sequence represents the 90kb region of the S. roseosporus daptomycin biosynthetic gene cluster
PT	PS	Sequence 90600 BP; 12671 A; 32312 C; 31571 G; 14046 T; 0 U; 0 Other;
PS	XX	Query Match 9.3%; Score 48.2; DB 6; Length 90600;
XX	XX	Best Local Similarity 52.8%; Pred. No. 3.1;
XX	XX	Matches 104; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
XX	QY	89 GAGCAGCGAGTCATCGCGCGCTCGATCTCGACGCTCGCGCGCAGCTCGATCTCGTCCCGG 148
XX	DB	78445 GAGCAGGAACCCGACGAGCGCGGCGGAGCGTCAAGTCGATCGAAGCGCGCCACCGCTGTCGTG 78386
XX	QY	149 CAGCAGCGGAGTCGCGCGCTCGATCTCGACGCTCGCGCGCAGCTCGATCTCGTCCCGCG 208
XX	DB	78385 CAGCAGCGTTCGATCTCGGCGCTTCTGCTCGCGCGCTCGAGCTTCGCGACGCTCGGTGTG 78326
XX	QY	209 CTCCACACACGAGGAAACTGGCTCGGTGGCAGCGGAGCCCCAGCCGCGGACAGTCTGTC 268
XX	DB	78325 CTCCACGCGCACCGCAACCTCGCGGGATGATCTGCAGCGGCGTGCAGCGGCGTGCAGCGGCTCGTA 78265
XX	QY	269 CAGGCTGGCGTCACCGA 285
XX	DB	78265 CCGGAAGCGCGCGCGCA 78249
XX	DE	RESULT 34
XX	DE	ADB68842
XX	DE	ID ADB68842 standard; DNA; 536 BP.
XX	AC	ADB68842;
XX	DT	04-DEC-2003 (first entry)
XX	DE	Minority luxI consensus sequence DNA 14.
XX	XX	quorum sensing; lux homologue; luxI; ds.
XX	OS	Unidentified.
XX	XX	WO2003057902-A2.
XX	PN	17-JUL-2003.
XX	PD	08-JAN-2003; 2003WO-US000479.
XX	PF	08-JAN-2002; 2002US-0346531P.
XX	PR	07-JAN-2003; 2003US-00338110.
XX	PA	(FRAU ) FRAUNHOFER USA INC.
XX	PI	Fuhrmann JJ, Romesser JA;







```
XX
DT 26-MAR-2002 (first entry)
DE Human herpesvirus 2 complete DNA genome.
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Human herpesvirus 2.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011372.
XX
PR 07-APR-2000; 2000US-0195680P.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
PI WPI; 2002-066308/09.
XX
DR Composition for oral delivery of vaccines, comprises expression vector
XX containing antigenic genomic sequence, bound to aggregated protein-
XX polycationic polymer conjugate or suspension.
XX
PS Disclosure; Page 90-132; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
XX bound to an aggregated protein-polycationic polymer conjugate or
XX suspension. The expression vector contains a promoter polynucleotide
XX sequence operatively linked to a polynucleotide sequence encoding an
XX antigen which is a fragment of a gene or genome associated with an
XX infectious disease, cancer and autoimmune disease such as rheumatoid
XX arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
XX consisting of bacterium, fungus, protozoa and virus such as human
XX immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX virus (HCV), influenza and respiratory syncytial virus (RSV), and
XX optionally comprising a nucleotide sequence encoding a cytokine (or a
XX cytokine expression vector), is useful for inducing an immune response
XX (systemic and/or mucosal) in an organism. The cytokine expression vector
XX contains a sequence for granulocyte macrophage-colony stimulating factor
XX (GM-CSF) or interleukin-12 (IL-12). The polynucleotide sequences encoding
XX the antigen and the cytokine are under transcriptional control of same or
XX different promoter polynucleotide sequences. The expression vector, as a
XX DNA vaccine is useful for treating a condition in an organism. The
XX present sequence is human herpesvirus 2 complete DNA genome related to
XX the invention
XX
SQ Sequence 154746 BP; 23003 A; 54218 C; 54701 G; 22824 T; 0 U; 0 Other;
XX
Query Match 9.2%; Score 48; DB 6; Length 154746;
Best Local Similarity 46.8%; Pred. No. 3.2;
Matches 233; Conservative 0; Mismatches 255; Indels 12; Gaps 2;
XX
QY 23 GTGCGCCGCGCCGCCATCAGCAACTGCCACTCCGCGGGGTGACCCAGGTCCACGAGTGTG 82
DB 129000 GTTGCCCGCCAGCGCGCGTGTGCGGGCCCGCAGAGCCGGTTGCCAGGCGCCGACGAG 129059
QY 83 GAGATTGAGCAGCCAGCTATCCCGCGCTGCGCGCGGATGCCGAAACACCGGCGCAGGAT 142
DB 129060 GCAGGACAGCCCGCGCTCGCGGACCACTCCGGGGGGGGCGCCCGCCCGCGGGCC 129119
QY 143 CAGGTGAGCAGCGGATGCGCGCTCGATCTCGACGTCGCGCGGCGAGTCTCGTTC 202
DB 129120 CCGGGCCAGGTCTCTCGCCCGGAGCGCGGAGTAGAGAT-----CACCACGCGACGTC 129173
QY 203 GCOCGGCTCCACACACAGGGGAAACTGGTCTGGTGTGACGCGGACGCCCGCGGACAG 262
```

```
Db 129174 CTCGGGTCCGGGCACCTGGGCGCATCCAGGCCGCGCGCGCGCGCGCGCGCGCGCG 129233
QY 263 CTGTCACAGGTGCGGTCTACGAGACCTTCGCGGGTTCGGGAGCGGAAACGCGCACGGCTCA 322
Db 129234 CAGCGGSCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129293
QY 323 GATCCCTGTAGTCGCTACGCTCAGTCCCGGTGTCGCCCTTGGCCCTGGGAGATAGCG 382
Db 129294 CGAGGCCAGCGCGCGCGGTCTGCAACATGAGGGCGCGCGCGCGCGCGCGCGCGCG 129353
QY 383 GTTCACAGCAGCGGACACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 442
Db 129354 GTGGTCCGTGAGCTCGGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129413
QY 443 TGACCA-----CGCGCTCGCGGGTTCGCGCGCGGTTCGCGCGAGATCCGTACCGCGCGCGCG 496
Db 129414 GGGCGCGCGGTGTGGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129473
QY 497 CCTCGGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
Db 129474 GCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129493
XX
RESULT 37
AAD25519/C
ID AAD25519 standard; DNA; 154746 BP.
XX
AC AAD25519;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human herpesvirus 2 complete DNA genome.
XX
KW Human herpesvirus 2; cytostatic; cancer; immunosuppressive; virucide;
KW antibacterial; fungicide; protozoicide; antirheumatic; antiinflammatory;
KW antiarthritic; rheumatoid arthritis; neuroprotective; multiple sclerosis;
KW immune response; vasotropic; vaccine; gene therapy; autoimmune disease;
KW vasculitis; ds.
XX
OS Human herpesvirus 2.
XX
PN WO200176643-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-US011372.
XX
PR 07-APR-2000; 2000US-0195680P.
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
PI Orson FM, Kinsey BM, Bhogal BS;
XX
PI WPI; 2002-066308/09.
XX
DR Composition for oral delivery of vaccines, comprises expression vector
XX containing antigenic genomic sequence, bound to aggregated protein-
XX polycationic polymer conjugate or suspension.
XX
PS Disclosure; Page 90-132; 145pp; English.
XX
CC The invention relates to a composition comprising an expression vector
XX bound to an aggregated protein-polycationic polymer conjugate or
XX suspension. The expression vector contains a promoter polynucleotide
XX sequence operatively linked to a polynucleotide sequence encoding an
XX antigen which is a fragment of a gene or genome associated with an
XX infectious disease, cancer and autoimmune disease such as rheumatoid
XX arthritis, vasculitis, and multiple sclerosis, pathogenic genomes
XX consisting of bacterium, fungus, protozoa and virus such as human
XX immunodeficiency virus (HIV), herpes simplex virus (HSV), hepatitis C
XX virus (HCV), influenza and respiratory syncytial virus (RSV), and
XX optionally comprising a nucleotide sequence encoding a cytokine (or a
```



XX  
SQ Sequence 4523 BP; 723 A; 1506 C; 1619 G; 675 T; 0 U; 0 Other;  
Query Match 9.2%; Score 47.6; DB 9; Length 4523;  
Best Local Similarity 53.2%; Pred. No. 5.1;  
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 325 TCCTGTGTCGATCGGTCAGTCCGGTCTCTCCCTTGGCTGGGAGATAGCGGT 384  
DB 3216 TCGCTGTAGGCCCGCCGAGGAGCGCGGCGGACCGGGCCCGGCGCGG 3157  
QY 385 TCACGAGGAGCGGCACACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 444  
DB 3156 GGGGAGCGGCCCGCCCTCCCGCGCTCCCGCGCGCGCGCGCGAGAGGGGG 3097  
QY 445 ACCAGCGGCTGCGGGGTCGGGGCGGTCGGCGAGATCCGTACCGCCCGGACCGCTCGGCC 504  
DB 3096 AACGGGGGGGACGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3037  
QY 505 AGCGCCGCGG 514  
DB 3036 GACGCGCGG 3027  
RESULT 40  
ADE60229/C  
ID ADE60229 standard; DNA; 4523 BP.  
XX  
AC ADE60229;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human gene M11167, SEQ ID NO 6135.  
XX  
KW Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; M11167.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a polynucleotide, a host cell  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates

QY 496 GCTCGGCGCAGCGC 509  
DB 877 ACCACACACAC 890  
RESULT 39  
ADE60227/C  
ID ADE60227 standard; DNA; 4523 BP.  
XX  
AC ADE60227;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human gene M11167, SEQ ID NO 6133.  
XX  
KW Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; M11167.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a polynucleotide, a host cell  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 4523 BP; 723 A; 1506 C; 1619 G; 675 T; 0 U; 0 Other;

Query Match 9.2%; Score 47.6; DB 9; Length 4523;  
Best Local Similarity 53.2%; Pred. No. 5.1;  
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 325 TCCTGTGTCAGTCGATCGGTCAGTCAGTCGCGGTCTCCCTTGGCGGAGGATAGCGGT 384  
DB 3216 TCGTGTCTAGGCG 3157

QY 385 TCACGACGAGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444  
DB 3156 GGGGACG 3097

QY 445 ACCAGCGGTCGCGGCGGTCGCGGCGGTCGCGGAGATCCGTACCGCCCGACCGCTCGGCC 504  
DB 3096 AACGGGGGGGACG 3037

QY 505 AGCGCGCGCG 514  
DB 3036 GACGCGCGCG 3027

RESULT 41  
ADE60231/c  
ID ADE60231 standard; DNA; 4523 BP.  
XX  
AC ADE60231;  
XX  
XX 29-JAN-2004 (first entry)  
XX Human gene M11167, SEQ ID NO 6137.  
XX  
XX Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX Homo sapiens.  
XX WO2003016475-A2.  
XX 27-FEB-2003.  
XX 14-AUG-2002; 2002WO-US025765.  
XX 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX (GEO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX Woolf C, D'urso D, Befort K, Costigan M;  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX

DR WPI, 2003-268312/26.  
DR GENBANK; M11167.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
CC specification) which encodes one of the polypeptides of the invention  
CC which is differentially expressed during pain. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 4523 BP; 723 A; 1506 C; 1619 G; 675 T; 0 U; 0 Other;

Query Match 9.2%; Score 47.6; DB 9; Length 4523;  
Best Local Similarity 53.2%; Pred. No. 5.1;  
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 325 TCCTGTGTCAGTCGATCGGTCAGTCAGTCGCGGTCTCCCTTGGCGGAGGATAGCGGT 384  
DB 3216 TCGTGTCTAGGCG 3157

QY 385 TCACGACGAGCGGACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 444  
DB 3156 GGGGACG 3097

QY 445 ACCAGCGGTCGCGGCGGTCGCGGCGGTCGCGGAGATCCGTACCGCCCGACCGCTCGGCC 504  
DB 3096 AACGGGGGGGACG 3037

QY 505 AGCGCGCGCG 514  
DB 3036 GACGCGCGCG 3027

RESULT 42  
ADE60233/c  
ID ADE60233 standard; DNA; 4523 BP.  
XX  
AC ADE60233;  
XX  
XX 29-JAN-2004 (first entry)  
XX Human gene M11167, SEQ ID NO 6139.  
XX Human; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.

XX OS Homo sapiens.  
XX FN WO2003016475-A2.  
XX PD 27-FEB-2003.  
XX PF 14-AUG-2002; 2002WO-US025765.  
XX PR 14-AUG-2001; 2001US-0312147P.  
XX PR 01-NOV-2001; 2001US-0346382P.  
XX PR 26-NOV-2001; 2001US-0333347P.  
XX PA (GEHO ) GEN HOSPITAL CORP.  
XX PA (FARB ) BAYER AG.  
XX PI Woolf C, D'urso D, Befort K, Costigan M;  
XX DR WPI; 2003-268312/26.  
XX DR GENBANK; M11167.  
XX PT New composition comprising two or more isolated polypeptides, useful for  
XX PT preparing a medicament for treating pain in an animal.  
XX PS Claim 1; Page; 1017pp; English.  
XX CC The invention discloses a composition comprising two or more isolated rat  
XX CC or human polynucleotides or a polynucleotide which represents a fragment,  
XX CC derivative or allelic variation of the nucleic acid sequence. Also  
XX CC claimed are a vector comprising the novel polynucleotide, a host cell  
XX CC comprising the vector, a method for identifying a nucleotide sequence  
XX CC which is differentially regulated in an animal subjected to pain and a  
XX CC kit to perform the method, an array, a method for identifying an agent  
XX CC that increases or decreases the expression of the polynucleotide sequence  
XX CC that is differentially expressed in neuronal tissue of a first animal  
XX CC subjected to pain, a method for identifying a compound which regulates  
XX CC the expression of a polynucleotide sequence which is differentially  
XX CC expressed in an animal subjected to pain, a method for identifying a  
XX CC compound that regulates the activity of one or more of the  
XX CC polynucleotides, a method for producing a pharmaceutical composition, a  
XX CC method for identifying a compound or small molecule that regulates the  
XX CC activity in an animal of one or more of the polypeptides given in the  
XX CC specification, a method for identifying a compound useful in treating  
XX CC pain and a pharmaceutical composition comprising the one or more  
XX CC polypeptides or their antibodies. The polynucleotide or the compound that  
XX CC modulates its activity is useful for preparing a medicament for treating  
XX CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
XX CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
XX CC specification) which encodes one of the polypeptides of the invention  
XX CC which is differentially expressed during pain. Note: The sequence data  
XX CC for this patent did not form part of the printed specification, but was  
XX CC obtained in electronic form directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 4523 BP; 723 A; 1506 C; 1619 G; 675 T; 0 U; 0 Other;

Query Match 9.2%; Score 47.6; DB 9; Length 4523;  
Best Local Similarity 53.2%; Pred. No. 5.1;  
Matches 101; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 325 TCCCTGTGAGTCGATCGCTCAGTGCCTGGTCTGCTCCCTTTGGCTGGAGATAGCGGT 384  
DB 3216 TCGCTGTAGTCCCGCCGAGCGAGGCGCGCGGACCGGCCCGCGGCGGACCCCGG 3157  
QY 385 TCACACAGCGGACACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444  
DB 3156 GGGGACGCGCGCGCGCCCTCGCGCCCTGCGCGCGCGCGCGCGCGCGCGGAGAGGGGG 3097  
QY 445 ACCACGGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGTGCGGCGGT 504  
DB 3096 AACGGGGGGGACGCGGCGGCGGCGGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3037

QY 505 AGCGCGCGCG 514  
DB 3036 GACGCGCGCG 3027

RESULT 43  
ADD45822/C  
ID ADD45822 standard; DNA; 4523 BP.  
XX AC ADD45822;  
XX DT 29-JAN-2004 (first entry)  
XX DE Human gene M11167, SEQ ID NO 11491.  
XX KW Human; ds; gene; pain; neuronal tissue; gene therapy;  
XX KW spinal segmental nerve injury; chronic constriction injury; CCI;  
XX KW spared nerve injury; SNI; Chung.  
XX OS Homo sapiens.  
XX PN WO2003016475-A2.  
XX PD 27-FEB-2003.  
XX PF 14-AUG-2002; 2002WO-US025765.  
XX PR 14-AUG-2001; 2001US-0312147P.  
XX PR 01-NOV-2001; 2001US-0346382P.  
XX PR 26-NOV-2001; 2001US-0333347P.  
XX PA (GEHO ) GEN HOSPITAL CORP.  
XX PA (FARB ) BAYER AG.  
XX PI Woolf C, D'urso D, Befort K, Costigan M;  
XX DR WPI; 2003-268312/26.  
XX DR GENBANK; M11167.  
XX PT New composition comprising two or more isolated polypeptides, useful for  
XX PT preparing a medicament for treating pain in an animal.  
XX PS Claim 1; Page; 1017pp; English.  
XX CC The invention discloses a composition comprising two or more isolated rat  
XX CC or human polynucleotides or a polynucleotide which represents a fragment,  
XX CC derivative or allelic variation of the nucleic acid sequence. Also  
XX CC claimed are a vector comprising the novel polynucleotide, a host cell  
XX CC comprising the vector, a method for identifying a nucleotide sequence  
XX CC which is differentially regulated in an animal subjected to pain and a  
XX CC kit to perform the method, an array, a method for identifying an agent  
XX CC that increases or decreases the expression of the polynucleotide sequence  
XX CC that is differentially expressed in neuronal tissue of a first animal  
XX CC subjected to pain, a method for identifying a compound which regulates  
XX CC the expression of a polynucleotide sequence which is differentially  
XX CC expressed in an animal subjected to pain, a method for identifying a  
XX CC compound that regulates the activity of one or more of the  
XX CC polynucleotides, a method for producing a pharmaceutical composition, a  
XX CC method for identifying a compound or small molecule that regulates the  
XX CC activity in an animal of one or more of the polypeptides given in the  
XX CC specification, a method for identifying a compound useful in treating  
XX CC pain and a pharmaceutical composition comprising the one or more  
XX CC polypeptides or their antibodies. The polynucleotide or the compound that  
XX CC modulates its activity is useful for preparing a medicament for treating  
XX CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene  
XX CC therapy). The sequence presented is a human DNA (shown in Table 2 of the  
XX CC specification) which encodes one of the polypeptides of the invention  
XX CC which is differentially expressed during pain. Note: The sequence data  
XX CC for this patent did not form part of the printed specification, but was  
XX CC obtained in electronic form directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences.

FT	/product= "Protein of ORF 11"	
FT	15880. .19035	CDS
FT	/*tag= l	
FT	/product= "Protein of ORF 12"	
FT	19032. .39713	CDS
FT	/*tag= m	
FT	/product= "Protein of ORF 13"	
FT	39713. .65800	CDS
FT	/*tag= n	
FT	/product= "Protein of ORF 14"	
FT	65826. .66530	CDS
FT	/*tag= o	
FT	/product= "Protein of ORF 15"	
FT	66546. .67370	CDS
FT	/*tag= p	
FT	/product= "Protein of ORF 16"	
FT	67384. .70059	CDS
FT	/*tag= q	
FT	/product= "Protein of ORF 17"	
FT	70099. .70662	CDS
FT	/*tag= r	
FT	/product= "Protein of ORF 18"	
FT	70659. .71906	CDS
FT	/*tag= s	
FT	/product= "Protein of ORF 19"	
FT	complement(73439. .71964)	CDS
FT	/*tag= t	
FT	/product= "Protein of ORF 20"	
FT	complement(74216. .73563)	CDS
FT	/*tag= u	
FT	/product= "Protein of ORF 21"	
FT	complement(75424. .74213)	CDS
FT	/*tag= v	
FT	/product= "Protein of ORF 22"	
FT	75535. .76464	CDS
FT	/*tag= w	
FT	/product= "Protein of ORF 23"	
FT	complement(78110. .76449)	CDS
FT	/*tag= x	
FT	/product= "Protein of ORF 24"	
FT	complement(79864. .78107)	CDS
FT	/*tag= y	
FT	/product= "Protein of ORF 25"	
FT	complement(81624. .79861)	CDS
FT	/*tag= z	
FT	/product= "Protein of ORF 26"	
FT	complement(81909. .81682)	CDS
FT	/*tag= aa	
FT	/product= "Protein of ORF 27"	
FT	complement(82345. .82062)	CDS
FT	/*tag= ab	
FT	/product= "Protein of ORF 28"	
FT	82587. .84446	CDS
FT	/*tag= ac	
FT	/product= "Protein of ORF 29"	
FT	84481. .85548	CDS
FT	/*tag= ad	
FT	/product= "Protein of ORF 30"	
FT	85556. .86845	CDS
FT	/*tag= ae	
FT	/product= "Protein of ORF 31"	
FT	87372. .86803	CDS
FT	/*tag= af	
FT	/product= "Protein of ORF 32"	
FT	87494. .88420	CDS
FT	/*tag= ag	
FT	/product= "Protein of ORF 33"	
XX		
PN	WO200231155-A2.	
XX		
PD	18-APR-2002.	
XX		
PP	15-OCT-2001; 2001WO-CA001462.	

XX 13-OCT-2000; 2000US-0239924P.  
PR 12-APR-2001; 2001US-0283296P.  
PR 24-JUL-2001; 2001US-00910813.  
XX  
XX  
XX (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
XX Farnet CM, Zazopoulos E, Staffa A;  
XX  
XX WPI; 2002-435445/46.  
DR P-PSDB; AAO22146, AAO22147, AAO22148, AAO22149, AAO22150, AAO22151,  
DR AAO22152, AAO22153, AAO22154, AAO22155, AAO22156, AAO22157, AAO22158,  
DR AAO22159, AAO22160, AAO22161, AAO22162, AAO22163, AAO22164, AAO22165,  
DR AAO22166, AAO22167, AAO22168, AAO22169, AAO22170, AAO22171, AAO22172,  
DR AAO22173, AAO22174, AAO22175, AAO22176, AAO22177, AAO22178.  
XX  
XX Novel isolated ramoplanin biosynthetic pathway polypeptide useful for  
XX chemically modifying biological molecule that is a substrate for a  
XX polypeptide encoded by a ramoplanin biosynthesis gene cluster.  
XX  
XX Disclosure; Page 87-135; 212pp; English.  
XX  
XX The invention relates to an isolated ramoplanin biosynthetic pathway  
XX polypeptide selected from a polypeptide of open reading frames (ORF) 1-  
XX 32. The isolated polypeptides are useful for chemically modifying a  
XX biological molecule that is a substrate for a polypeptide encoded by a  
XX ramoplanin biosynthesis gene cluster, by contacting the biological  
XX molecule with the isolated polypeptide, where the polypeptide chemically  
XX modifies the biological molecule. The method comprises contacting the  
XX biological molecule with at least two different polypeptides encoded by  
XX ramoplanin ORFs 1-31. The polypeptides are useful for directing the  
XX biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated  
XX gene cluster comprising the ORFs is useful as a substrate for  
XX bioengineering of antibiotic structures. An isolated polypeptide or its  
XX encoding nucleic acid sequence is useful for generating derivatives of  
XX ramoplanin, for improving production or for producing variants of other  
XX antibiotics of the peptide class. The isolated polypeptides are useful  
XX for synthesis of ramoplanin in vivo or in vitro, as an acylation domain  
XX in conjunction with other peptide synthetase modules and allowing the  
XX incorporation of Thr into a peptide antibiotic precursor, for modifying  
XX fatty acid structure and/or enhancing fatty acid incorporation into the  
XX peptide antibiotic structure, for production of an hydroxyphenylglycine  
XX (HPG)-containing peptide antibiotic, for enhancing secretion of  
XX ramoplanin or its variants and derivatives, for enhancing uptake of  
XX precursors for ramoplanin biosynthesis, for enhancing production of  
XX ramoplanin products or its variants or derivatives, to chlorinate HPG of  
XX a peptide antibiotic precursor, and for designing specific nucleotide  
XX probes and primers for identifying and isolating putative lipopeptide  
XX -producing microorganisms. This polynucleotide sequence represents the  
XX 88421nt genomic DNA of a ramoplanin producing Actinoplanes sp.  
XX microorganism of the invention  
XX  
XX Sequence 88421 BP; 10639 A; 32391 C; 32663 G; 12728 T; 0 U; 0 Other;  
XX  
XX Query Match 9.1%; Score 47.4; DB 6; Length 88421;  
XX Best Local Similarity 49.8%; Pred. No. 4.3;  
XX Matches 120; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
XX  
XX 43 CGAAGTCACCTCGCGCGGGTGACAGGTGACAGGTGGAGTTGACGACGACGTCAC 102  
XX 59630 CGACCGCCGTGCGATCGGGATGTCGGCGCGCGCGCGGACGACGCGTCGCGA 59571  
XX  
XX 103 TCGCGCCCTGCGCGGATGCGGAAACACCGGCGCAGGATACGTCGACGCGGATGC 162  
XX 59570 AGCGCGCTGACGACCATGACACGGTCACGCTTCGACCGGCGGCGGCGGCGG 59511  
XX  
XX 163 GCGCCCTGATTCGACGCTGCGCGGAGTTCGATTCGTCGCGCGGCTCCACACAGGG 222  
XX 59510 GCGCGTGCACCTCGCGCGGACCGCGCACGGGACCTCGTGTCTCCGCGTGGCTCGG 59451  
XX  
XX 223 GAAACTGGCTCGTGGCAGCGGACCGCGGCGGACGAGTCGTCAGGCTGGGTCAC 282  
XX 59450 CCGGGCGCGCGCGTGGCGCGGAGGTCCAGTCTCTCCGTTAGCGGTGACGCGTCC 59391

QY 283 G 283  
DB 59390 G 59390

## RESULT 45

ABK32842/C  
ID ABK32842 standard; DNA; 2307 BP.  
XX  
XX AC ABK32842;  
XX  
XX 23-APR-2002 (first entry)  
DT  
DE DNA encoding human homologue of MPT1 antifungal target.  
XX  
XX antifungal; fungal gene transcription; RRC34; POP3; TPA2; NAB2; MPT1;  
XX MTR2; BOS1; POL30; RSA2; SQT1; MTW1; TFB1; SPC99; BFR2; RNAL; GCD7; SK16;  
XX NP1; LCP5; NCE103; EOC1; ORC2; CNS1; YPD1; TIM10; SRB4; Yeast; fungus;  
XX ds; gene.  
XX  
XX Homo sapiens.  
XX  
XX WO200202055-A2.  
XX  
XX 10-JAN-2002.  
XX  
XX 28-JUN-2001; 2001WO-US020592.  
XX  
XX 29-JUN-2000; 2000US-0215164P.  
XX 10-AUG-2000; 2000US-0224457P.  
XX  
XX (ANAD-) ANADYS PHARM INC.  
XX  
XX Moore J, Buurman ET, Desilva T, Harris S, Komarnitsky S;  
XX Mendillo M, Moore D, McCoy M, Sanderson K, Haq T, Zhu S, Long P;  
XX Davidov E, Thompson CM;  
XX  
XX WPI; 2002-147962/19.  
XX P-PSDB; AAU82954.  
XX  
XX Screening candidate antifungal compound for interaction with essential  
XX protein, modulation of essential protein activity, binding to essential  
XX protein, by contacting protein with test compound and determining  
XX effects.  
XX  
XX Disclosure; Fig 80; 522pp; English.  
XX  
XX The invention describes a method of screening a candidate antifungal  
XX compound for interaction with essential proteins (EP) or for modulation  
XX of EP activity e.g. fungal gene transcription. The proteins tested in the  
XX invention include RPC34, POP3, TPA2, NAB2, MPT1, MTR2, BOS1, POL30, RSA2,  
XX SQT1, MTW1, TFB1, SPC98, BFR2, RNAL, GCD7, SK16, NIP1, LCP5, NCE103,  
XX EOC1, ORC2, CNS1, YPD1, TIM10 and SRB4 from S. cerevisiae, C. albicans  
XX or more test compounds and determining the effects on the growth or  
XX viability of the culture of cells which preferably comprises fungal cells  
XX or yeast cells. Preferably the identified compounds interact with, or  
XX modulate (preferably inhibit) activity of C. albicans EP. The inhibitor  
XX compounds identified by the method are useful for preventing or  
XX inhibiting fungal, particularly C. albicans growth in culture or in a  
XX mammal. The antifungal agents interact with essential fungal elements  
XX that can be used to treat fungal infection by preventing the growth and  
XX preferentially killing the fungi, but does not inhibit the biological  
XX activity of mammalian homologues. This sequence encodes a target protein  
XX used to test the antifungal compounds, described in the method of the  
XX invention  
XX  
XX Sequence 2307 BP; 359 A; 969 C; 724 G; 255 T; 0 U; 0 Other;  
XX

Query Match 9.1%; Score 47.2; DB 6; Length 2307;  
Best Local Similarity 44.6%; Pred. No. 6.4;  
Matches 232; Conservative 0; Mismatches 283; Indels 5; Gaps 1;



QY 2 GTGACGACGAGACGAGCGGTGTCGCGCCGCGCCGCGCCATCAGAACTGCCACTCCGGCGG 61  
Db 1269 GGGCGTCCGGGACAGGCTCTGGGTCACTGCGCGCGCGCGCCCTTTGGGGCAGCCCGGTGGG 1210  
QY 62 GGTGACAGAGTGCACAGAGTGGAGTTGAGCAGCAGCTCATCGCGCGCTGGCGCGGAT 121  
Db 1209 GGTCCCGGGCGCGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1150  
QY 122 GCCGAAACACCGCGCCAGGATCAGCTGACAGCGCGGATGCGCGCGCTCGATCTTCGACCGT 181  
Db 1149 CCGGATGACCATGCTGCGCGCGCGCTGCTGGCGCGCGCGCTGGCGCGCGCGCGCGCG 1090  
QY 182 CGCGCGAGTGCATCTGTCGCGCGCGCGCTCCACACAGCGGGAATCTGCTCGTGGCGAG 241  
Db 1089 CG 1030  
QY 242 CGGACG 301  
Db 1029 CGGACG 970  
QY 302 AGCGGAAACCG 356  
Db 969 GCG 910  
QY 357 GTCCCGCTTGGCG 416  
Db 909 ATACCTTGGCG 850  
QY 417 GGGCGCGGTTGACCGCGATCGCTGATGACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476  
Db 849 GCG 790  
QY 477 GATCGTACCG 516  
Db 789 GCG 750  
RESULT 46  
ID ABX72218/c  
XX ABX72218 standard; cDNA; 5571 BP.  
AC ABX72218,  
DT (first entry)  
XX 03-JUN-2003  
XX Human NOVX polynucleotide #49.  
XX Human; NOVX; gene; ss; metabolic disorder; cardiomyopathy; diabetes; ASD;  
KW hypertension; congenital heart defect; aortic stenosis; valve disease;  
KW atrial septal defect; atrioventricular canal defect; ductus arteriosus;  
KW pulmonary stenosis; subaortic stenosis; ventricular septal defect; VSD;  
KW tuberculous sclerosis; scleroderma; atherosclerosis; infectious disease;  
KW obesity; anorexia; neurodegenerative disorder; Alzheimer's disease;  
KW Parkinson's disease; immune disorder; haematopoietic disorder;  
KW haemophilia; hypercoagulation; Crohn's disease; cancer.  
XX Homo sapiens.  
OS  
XX  
XX WO20281498-A2.  
XX  
XX 17-OCT-2002.  
XX  
XX 03-APR-2002; 2002WO-US010780.  
XX  
XX 03-APR-2001; 2001US-0281086P.  
PR 03-APR-2001; 2001US-0281136P.  
PR 05-APR-2001; 2001US-0281863P.  
PR 05-APR-2001; 2001US-0281906P.  
PR 06-APR-2001; 2001US-0282020P.  
PR 10-APR-2001; 2001US-0282930P.  
PR 10-APR-2001; 2001US-0282934P.  
PR 12-APR-2001; 2001US-0283512P.

PR 13-APR-2001; 2001US-0283710P.  
PR 17-APR-2001; 2001US-0284234P.  
PR 19-APR-2001; 2001US-0285325P.  
PR 20-APR-2001; 2001US-0285381P.  
PR 20-APR-2001; 2001US-0285609P.  
PR 23-APR-2001; 2001US-0285748P.  
PR 23-APR-2001; 2001US-0285890P.  
PR 24-APR-2001; 2001US-0286068P.  
PR 25-APR-2001; 2001US-0287213P.  
PR 27-APR-2001; 2001US-0287252P.  
PR 02-MAY-2001; 2001US-0288257P.  
PR 29-MAY-2001; 2001US-0294164P.  
PR 30-MAY-2001; 2001US-0294484P.  
PR 18-JUN-2001; 2001US-0298952P.  
PR 19-JUN-2001; 2001US-0299237P.  
PR 19-JUN-2001; 2001US-0299276P.  
PR 12-SEP-2001; 2001US-0318750P.  
PR 25-SEP-2001; 2001US-0324800P.  
PR 25-SEP-2001; 2001US-0324802P.  
PR 27-SEP-2001; 2001US-0325684P.  
PR 17-OCT-2001; 2001US-0330143P.  
PR 14-NOV-2001; 2001US-0332131P.  
PR 14-NOV-2001; 2001US-0332240P.  
PR 14-NOV-2001; 2001US-0332775P.  
PR 21-NOV-2001; 2001US-0332115P.  
PR 04-DEC-2001; 2001US-0337621P.  
PR 03-JAN-2002; 2002US-0345783P.  
PR 16-JAN-2002; 2002US-0350251P.  
PR 02-APR-2002; 2002US-00114270.  
XX (CURA-) CURAGEN CORP.  
XX  
PA  
PI Guo X, Kekuda R, Miller CE, Malyankar UM, Spytek KA;  
PI Patturajan M, Liu X, Gusev VV, Li L, Vernet CAM, Zerhusen BD;  
PI Gorman L, Shenoy SG, Pena CE, Smithson G, Burgess CE, Gerlach V;  
PI Padigaru M, Shinketsa RA, Gangolli EA, Taupier RJ, Casman SJ, Ji W;  
PI Anderson DW, Leite MW, Rastelli L, Edinger SR, Stone DJ;  
PI Macdougall JR, Rothenberg ME, Mazur A, Millet I, Peyman JA;  
PI Ellerman K;  
XX  
DR WPI: 2003-046858/04.  
DR P-PSDB; ABUS4590.  
XX  
PT New isolated NOVX polypeptide useful for treating atherosclerosis,  
PT metabolic disorders, diabetes, obesity, infectious disease, anorexia,  
PT neurodegenerative disorders, Alzheimer's disease and cancer.  
XX  
PS Claim 17; Page 196-197; 666pp; English.  
XX  
CC The invention relates to human polypeptides, termed NOVX, and the  
CC polynucleotides encoding them. The polypeptides and polynucleotides are  
CC useful for diagnosing disease, and screening for potential therapeutic  
CC agents. The sequences are useful for treating metabolic disorders,  
CC cardiomyopathy, diabetes, hypertension, congenital heart defects, aortic  
CC stenosis, atrial septal defect (ASD), atrioventricular canal defect,  
CC ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular  
CC septal defect (VSD), valve diseases, tuberculous sclerosis, scleroderma,  
CC atherosclerosis, obesity, infectious disease, anorexia, neurodegenerative  
CC disorders, Alzheimer's disease, Parkinson's disease, immune disorders,  
CC haematopoietic disorders, haemophilia, hypercoagulation, Crohn's disease,  
CC and cancer. Sequences ABX72170-ABX72275 represent human NOVX  
CC polynucleotides of the invention  
XX  
SQ Sequence 5571 BP; 924 A; 2087 C; 1813 G; 747 T; 0 U; 0 Other;  
Query Match 9.1%; Score 47.2; DB 7; Length 5571;  
Best Local Similarity 47.1%; Pred. NO. 5.9;  
Matches 178; Conservative 0; Mismatches 198; Indels 2; Gaps 1;  
QY 8 CACGAACGACCGGTGCTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 67  
Db 390 CACGGCG 331



QY 68 CAGTCGACGAGTGGAGTTGAGCAGCCAGCTCATCCG--CGCTCGCCGCGCATCCG 125  
Dd |||||  
330 GGCACGACGAGCGCTGGCCCGCGGCGCGTCCATCCCGCGCGCGCGCGCGCGG 271  
QY 126 AAACACGGGGCAGCATACGTCGACGACGCGGATGCGGCTCGATCTCAGCGTGGC 185  
Dd |||||  
270 GCGCTGCGCGCGGGAACGAGAGCCCGGGGACTCCCGCGGGCGCGCCCGCGCC 211  
QY 186 CGCAGCTCGATCTCGTCGCGCGGCTCCACACACAGGGGAACTGGCTCGTGGCAGCGGC 245  
Dd |||||  
210 CTTCCCCCGCGAGCCCGTCCGCGCGCGTCCGCGGGAGGGCGCGGGCGCGG 151  
QY 246 AGCCCCAGCGGAGCAGCTCGTCAGGCTGGGTCACGAGCTCTCCGGGGTCCGAGCG 305  
Dd |||||  
150 GCGCGGGCGCGCGCGCGGAGCAGCGGCTCCGCGGGGCTCCGAGGGGGCGCGCGG 91  
QY 306 GAAACGCGCAGCGCTCAGATCCCTCTCAGTGGCACTCGGCTCAGTCCCGCTT 365  
Dd |||||  
90 GCGCGCGCGCGCTCAGTGCATCGCGCGCGCTCAGTCCGCGCTTCCCGCGG 31  
QY 366 GGCCTGGGAGGATAGCG 383  
Dd |||||  
30 GCGCGGGCGCGCGCGG 13

## RESULT 47

ACA42277/c

ID ACA42277 standard; DNA; 1611 BP.

XX AC

XX ACA42277;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #23934.

XX KW

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX KW drug design; gene.

XX OS

XX OS Pseudomonas aeruginosa.

XX PN

XX PN WO200277183-A2.

XX PD

XX PD 03-OCT-2002.

XX PF

XX PF 21-MAR-2002; 2002WO-US09107.

XX PR

XX PR 21-MAR-2001; 2001US-00815242.

XX PR

XX PR 06-SEP-2001; 2001US-00948993.

XX PR

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR

XX PR 08-FEB-2002; 2002US-00072851.

XX PR

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA

XX PA (BLIT-) ELITRA PHARM INC.

XX PI

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI

XX PI Wall D, Irawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX PI

XX PI WPI; 2003-029926/02.

XX DR

XX DR P-PSDB; ABU38407.

XX XX

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS

XX PS Claim 14; SEQ ID NO 30147; 1766pp; English.

XX CC

XX CC The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC on a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 1611 BP; 301 A; 571 C; 494 G; 245 T; 0 U; 0 Other;

## Query Match

Best Local Similarity 9.0%; Score 47; DB 7; Length 1611;

Matches 132; Conservative 0; Mismatches 115; Indels 6; Gaps 1;

QY 65 GACGAGTTCGACGAGTGGGAGTTGAGCAGCCAGCTCATCCGCGCTCGCGCGCATGCC 124

Dd |||||

660 GACCAGTTGAGTTCGTTGGGTGATCAGCAGCGGACATCCCGAGCGTCCGTCAGGTC 601

QY 125 GAAACACCGGGCCAGGATCAGTCGAGCAGCGGCGATCGCGCTCGATCTCAGCGTCGG 184

Dd |||||

600 CTTGAGCAGATCGAGGATCTTCAACTGGACGGTGA-----CGTCAGCGCGGTGTCGG 547

QY 185 CCGCAGCTCGATCTCGTCGCGCGCTCCACACAGCGGAAACTCGTGGTGGCAGCGG 244

Dd |||||

546 TTCTGCGCGGATCAGCAGTTGCGGCTCGCAGCGCAGGCGCATCATCAGCGGCTG 487

QY 245 CAGCCCCAGCGGGAGCAGCTCGTCCAGGTTGGGCTCAGGACCTCTCGCGGGTTCGGGAGC 304

Dd |||||

486 CCGCTGCGCGCGGAGAGTTGCTGTGATAGGCTTTTCAGCGGCTTGGCGCGCTCGGGGAT 427

QY 305 GGAACGCGCAGC 317

Dd |||||

426 GCGACCCAGCTCG 414

## RESULT 48

AAA30798/c

ID AAA30798 standard; DNA; 1835 BP.

XX AC

XX AAA30798;

XX XX

XX DT 29-AUG-2000 (first entry)

XX XX

XX DE DNA encoding maize hydroperoxide lyase (HPL).

XX KW

XX KW Hydroperoxide lyase; HPL; maize; volatile aldehyde; oxylipin pathway;

XX KW disease resistance; fungal infection; flavour; aroma; transgenic plant;

XX KW ds.

XX XX

XX OS

XX OS Zea mays.

XX FH

XX FH Key Location/Qualifiers

XX FT

XX FT 115..1624

XX FT

XX FT /\*cag= a

XX FT /product= "Maize hydroperoxide lyase"

XX XX



```
QY 451 GCGTCGGGGTCGGGGCGGTGGGCGAGATCCGTACCGCCCGGACCGCTCGGCGCAGCGCC 510
Db 173 CGCGCCCGCGCAGCGCGCGCCAGCGCGCGCTGCTCGCCCCAGAGCGGATCGCCAGCGTC 114
QY 511 GCGCGGATC 519
Db 113 TCCGCGCTTC 105

RESULT 50
AAI19682_39
Continuation (40 of 45) of AAI19682 from base 3900001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI19682 Accession AAI19682
WP Fragment Name Begin End
WP AAI19682_00 1 110000
WP AAI19682_01 100001 210000
WP AAI19682_02 200001 310000
WP AAI19682_03 300001 410000
WP AAI19682_04 400001 510000
WP AAI19682_05 500001 610000
WP AAI19682_06 600001 710000
WP AAI19682_07 700001 810000
WP AAI19682_08 800001 910000
WP AAI19682_09 900001 1010000
WP AAI19682_10 1000001 1110000
WP AAI19682_11 1100001 1210000
WP AAI19682_12 1200001 1310000
WP AAI19682_13 1300001 1410000
WP AAI19682_14 1400001 1510000
WP AAI19682_15 1500001 1610000
WP AAI19682_16 1600001 1710000
WP AAI19682_17 1700001 1810000
WP AAI19682_18 1800001 1910000
WP AAI19682_19 1900001 2010000
WP AAI19682_20 2000001 2110000
WP AAI19682_21 2100001 2210000
WP AAI19682_22 2200001 2310000
WP AAI19682_23 2300001 2410000
WP AAI19682_24 2400001 2510000
WP AAI19682_25 2500001 2610000
WP AAI19682_26 2600001 2710000
WP AAI19682_27 2700001 2810000
WP AAI19682_28 2800001 2910000
WP AAI19682_29 2900001 3010000
WP AAI19682_30 3000001 3110000
WP AAI19682_31 3100001 3210000
WP AAI19682_32 3200001 3310000
WP AAI19682_33 3300001 3410000
WP AAI19682_34 3400001 3510000
WP AAI19682_35 3500001 3610000
WP AAI19682_36 3600001 3710000
WP AAI19682_37 3700001 3810000
WP AAI19682_38 3800001 3910000
WP AAI19682_39 3900001 4010000
WP AAI19682_40 4000001 4110000
WP AAI19682_41 4100001 4210000
WP AAI19682_42 4200001 4310000
WP AAI19682_43 4300001 4410000
WP AAI19682_44 4400001 4411529

Query Match 9.0%; Score 47; DB 4; Length 110000;
Best Local Similarity 45.3%; Pred. No. 4.9;
Matches 208; Conservative 0; Mismatches 250; Indels 1; Gaps 1;
QY 57 GCGCGGGTGAACAGATCGACGAGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTGGCGCC 116
Db 31301 GCGGTGATCAATGCGCCACCGCGCTGCTGGGGCGCCGTTGATCGGTGACGGGCC 31360
QY 117 GCGATGCGGAACACCGGCGCAGGATCAGTGCAGCAGCGGAGTCGGCGCCTCGATCTCG 176
Db 31361 AATGCGACACTCCGCGCGGGCCCGCGCGGGCTGCTGCTTCGGAACGCGGG 31420
```

```
QY 177 ACGGTCGGCCGAGCTCGATCTCGTCGCCCGGCTCCCAACACAGGGGAACCTGGTTCGGT 236
Db 31421 GCGGGGCGAGCCGGGGCGCCCGCCAGGCCCGCGGGCCTTGGCGGCCCGCGGATTTGTGG 31480
QY 237 GGCAGCGGCGAGCCGCCAGCCCGGACAGCTCGTCCAGCTGGCGTCAAGACCTCTCGCGGG 296
Db 31481 GGCACGCGGGGCGCGCGGGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 31540
QY 297 TCGGAGCGGAAAACGCGCACCGCTCAGATCCCTGTCAGTTCGATCGGCTCAGTCCGGT 356
Db 31541 GCGGTGGGTGGCTGTTTCGGGGTTGGCGGGCGCGGGGTGTCGGTGGGGCCGGTGGCGG 31600
QY 357 GTCCCCCTTGGCCTTGGGAGGATAGCGGTTCAAGACGAGCGGACCAACGCGCGGGCGG 416
Db 31601 ACCGGCGGGCGGGCGGGCGGCCCGGTGTTGATCTGGGGCGGGCGGGCGGGCGGGTGC 31660
QY 417 GGGGGCGGTTTCAGCCGATCCGCTCGATGACACAGCGGCTCGGGGTTCGGGGCGGTTC 476
Db 31661 GGTGGGGCGGTTGGCGGCGACCGCGCGGG-CCGGCGGGCGCGCGGAGCTGCTGTTCGG 31719
QY 477 GATCCGTACCGCCCGGACCGCCTTCGGCCAGCGCGCGCGG 515
Db 31720 CGGCGGTGGGGTGGGGCGGGCACCGACGCGGGCGCGG 31758
```

Search completed: June 27, 2004, 18:18:40  
Job time : 346.02 secs



Micromonosporineae; Micromonosporaceae; Micromonospora.

REFERENCE

1 Hosted,T.J., Horan,A.C. and Wang,T.X.  
Everminomycin biosynthetic genes  
Patent: WO 0151639-A 1 19-JUL-2001;  
Schering Corporation (US)

FEATURES

source

1. 109519  
/organism="Micromonospora carbonacea"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:47853"

ORIGIN

Query Match 100.0%; Score 520; DB 6; Length 109519;  
Best Local Similarity 100.0%; Pred. No. 7.3e-68;  
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTGACAGCAACGACCGGTGGTTCGCGCCGGCGCCATCAGAACTGCCTCCGGCG 60  
Db 109000 GGTGACAGCAACGACCGGTGGTTCGCGCCGGCGCCATCAGAACTGCCTCCGGCG 109059  
QY 61 GGGTGAACAGGTGCACAGGTGGAGTTGAGCAGCCAGCTCATCGCGCGCTGCGCGGCA 120  
Db 109060 GGGTGAACAGGTGCACAGGTGGAGTTGAGCAGCCAGCTCATCGCGCGCTGCGCGGCA 109119  
QY 121 TGCCGAACACCGGGCCAGGATCACGTGCAGCAGCGCGATCGCGCTCGATCTCGACGG 180  
Db 109120 TGCCGAACACCGGGCCAGGATCACGTGCAGCAGCGCGATCGCGCTCGATCTCGACGG 109179  
QY 181 TCGCCCGCAGCTCGATCTCGTCCCGGCTCCACACACGGGAAACTGCTCGTGGCA 240  
Db 109180 TCGCCCGCAGCTCGATCTCGTCCCGGCTCCACACACGGGAAACTGCTCGTGGCA 109239  
QY 241 CGGCGACCCCGAGCGGAGCAGCTGTCAGGCTGGGTTCAGGACCTCTCGCGGTTCGG 300  
Db 109240 CGGCGACCCCGAGCGGAGCAGCTGTCAGGCTGGGTTCAGGACCTCTCGCGGTTCGG 109299  
QY 301 GAGCGGAACCGCACCGGCTCAGATCCCTGTGATGTCATCGGCTCAGTCCCGGTCCTCC 360  
Db 109300 GAGCGGAACCGCACCGGCTCAGATCCCTGTGATGTCATCGGCTCAGTCCCGGTCCTCC 109359  
QY 361 CCCTTGCGCTGGAGATAGCGTTACAGCAGCGGACACACGCGGGGGGGGGGG 420  
Db 109360 CCCTTGCGCTGGAGATAGCGTTACAGCAGCGGACACACGCGGGGGGGGGGG 109419  
QY 421 GCGGTTGAGCGATCCCGCTCGATACACGAGCGGTTCGCGGCTCGGCGGATC 480  
Db 109420 GCGGTTGAGCGATCCCGCTCGATACACGAGCGGTTCGCGGCTCGGCGGATC 109479  
QY 481 GGTACCGCGGACCGCTCGCGCAGCGCGCGCGGATCC 520  
Db 109480 GGTACCGCGGACCGCTCGCGCAGCGCGCGCGGATCC 109519

RESULT 2

AX196102/c

LOCUS

DEFINITION

AX196102

AX196102

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1

AX196102.1









```

Db 7759 AGAGGGGCTGATGGCAGAGTCCACGTCGAGCAGGAGCCCTGATCCGGCTGATCAGCAGCA 7700
Qy 181 TCGGCGGACGCTGATCTGCTGCGCCGCTCCACACACAGGAGAACTCGCTCGGTGGCA 240
Db 7699 TCGCCAGCAGGAGTGGCCCGCCCGAGGCGGAGAGTCTGCTCCACCCCGAGCCCGGCA 7640
Qy 241 GCGGACAGCCCGGAGCAGTCTGTCAGAGTGGCGTCCAGGACCTCTCGCGGGTGG 300
Db 7639 CCGGACGACCTCGGCGAAACACCGGACAGATCTCTCGCGGAGCGCGGGCGCC 7580
Qy 301 GAGCGGAAACGCGACGGCTCAGATCCCTGTCAGTCGCTCAGTCGCGGTGCTCC 360
Db 7579 GCCTCCCGCGCAGCGCGCGGAACTCCGCGCACCGGCGCGGCTCCAGCTTGC 7520
Qy 361 CCCTGGCGCTGGAGGATACGGTTACAGCAGGAGCGGACACCGGGCGG 410
Db 7519 CGTTGGCGCTACGGCAACCCGTTCTCCAGCACCACCGCGCGGCGG 7470

RESULT 7
AX803750/c
LOCUS AX803750 61944 bp DNA linear PAT 25-NOV-2003
DEFINITION Sequence: 34 from Patent WO03060127.
ACCESSION AX803750
VERSION AX803750.1 GI:38520921
KEYWORDS Streptomyces refuineus
SOURCE Streptomyces refuineus
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Farnet, C.M., Staffs, A. and Zazopoulos, E.
TITLE Genes and proteins involved in the biosynthesis of lipopeptides
JOURNAL Patent: WO 03060127-A 34 24-JUL-2003; Ecopia Biosciences Inc. (CA)
FEATURES
    source 1..61944
    /organism="Streptomyces refuineus"
    /mol_type="unassigned DNA"
    /db_xref="taxon:223296"

ORIGIN
    Query Match 11.5%; Score 59.6; DB 6; Length 61944;
    Best Local Similarity 46.6%; Pred. No. 4.1;
    Matches 191; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

Qy 1 GGTGAGCAGCAACGACGCGGTGTCGCCCGCGCGCCCATCAGAACTGCACCTCGGCG 60
Db 38215 GCTGCTGGCGGTAGGAGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38156
Qy 61 GGGTGACAGGCTGACACGAGTGGGAGTTGAGCAGCAGCTCATCGCGCGCTGCGCGGCA 120
Db 38155 CGGGCGCGCTGCGGCGCGCTGCGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCG 38096
Qy 121 TCGCGAAACACCGGCGCGGAGTACGTCAGCAGCGGCGGAGTGGCGGCTCGATCTCAGCG 180
Db 38095 AGAGGGCGCTCGATGCGCGGAGTCCAGCTGAGCAGGCGGCGGCGGCGGCGGCGG 38036
Qy 181 TCGCGCGGCTCGATCTCGTCCGCGCGGCTCCACACACAGGAGAACTGCTCGGTGGCA 240
Db 38035 TCGCGCAGGAGGATGGCGCGCGCGGCGGAGGAGAAAGTGTGTCTCCACCGCGCGGCGGCA 37976
Qy 241 CGCGCAGCGCGCGGAGCAGCTGCTCGAGGCTGGGCTCAGCGGCTCAGCGGCTCGG 300
Db 37975 CCGCGCAGCCTCGGCGGAAACACCGGACAGATCTCTCGCGGCGGCGGCGGCGGCGG 37916
Qy 301 GAGCGGAAACGCGCAGCTCAGATCTCTGTCAGTCCGCTCAGTCCGCTCGCTCC 360
Db 37915 GCCCAGCGCGCGCGCGGCGGAGTCCGCGCACCGGCGGCGGCGGCGGCGGCTGCG 37856
Qy 361 CCCTGGCGCTGGAGGATAGCGGTTACAGCAGGCGGCGGCGGCGGCGGCGGCGG 410

```

```

Db 37855 CGTTGGCGCTCAGCGCAACCGGCTCTCCAGCACCACCGCGCGCGGCGGCGG 37806

RESULT 8
SC0939118/c
LOCUS SC0939118 303550 bp DNA linear BCT 11-FEB-2003
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 15/29.
ACCESSION AL3939118 AL049826 AL079308 AL096823 AL118514 AL118515
AL356612 AL357152 AL392177 AL583944 AL583945 AL596251 AL645882
AL939118.1 GI:24413861
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, I., Murphy, L., Oliver, K., O'Neill, S., Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.
TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 303550)
AUTHORS Bentley, S.D.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sds@sanger.ac.uk
COMMENT On or before Oct 26, 2002 this sequence version replaced
    gi:20520812, gi:4808374, gi:5102782, gi:5457267, gi:5918466,
    gi:5918500, gi:8052426, gi:8247645, gi:10178355, gi:13122127,
    gi:13122180, gi:15021246;
    Location/Qualifiers
        1..303550
        /organism="Streptomyces coelicolor A3(2)"
        /mol_type="genomic DNA"
        /strain="A3(2)"
        /db_xref="taxon:100236"
        424..1296
        /gene="SCO3843"
        /note="SCO3843"
        /note="SCH69.13, unknown, len: 290 aa; similar to the N-terminal region of TR:P71590 (EMBL:Z80233) hypothetical protein from Mycobacterium tuberculosis (527 aa) fasta scores; opt: 268, z-score: 221.4, E(): 5.3e-05, (28.1% identity in 217 aa overlap)."
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="CAB45210.1"
        /db_xref="GI:5102795"
        /db_xref="GCA:Q9XA21"
        /db_xref="SPTREMBL:Q9XA21"
        /translation="MWLVKKEQRLGLVNGTFAKFKSEVQPVIEIAGALQRECDNNA
        TWRDRVTWDFVLESTPDLSPYQGLDELQAGVMDYAKQRYTFMGPIKV
        HLEKADLDLTGLYRVSRTSLASTSHQGGPAAAPAPPPAGPGGYPGPAAPPA
        GLEKADLDLTGLYRVSRTSLASTSHQGGPAAAPAPPPAGPGGYPGPAAPPA
        GLEKADLDLTGLYRVSRTSLASTSHQGGPAAAPAPPPAGPGGYPGPAAPPA
        MGRSTEADVRIDDPGVSRHCEIRTGTTSTIQDLGSTNGIVVDGQHTTRATLRDGR
        IVGSTTVIYRQAEQ"
        1307..1825
        /gene="SCO3844"
        /note="synonym: SCH69.14"
        1307..1825

```

```

/gene="SCO3844"
/note="SCH69.14, putative secreted protein, len: 172 aa;
similar to TR:P71589 (EMBL:Z80233) hypothetical protein
from Mycobacterium tuberculosis (155 aa) fasta scores;
opt: 399, z-score: 442.8, E(): 2.5e-17, (46.2% identity in
171 aa overlap). Contains possible N-terminal region
signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB45211.1"
/db_xref="GI:5102796"
/db_xref="SPTREMBL:O9XA20"
/translacion="MSLTLTVMELGLVAVIVAVIQRSDLEGTGTORGARR
DARPQATRGQAGAPFPQSCGGRGRGATKLVTEGTLRTVALQGGTYILG
RAHSTVLDLDDYDASSRHARIYPQNGQWIVDGLSTGYLDRARLTTPTPGPGAP
IRICKTVIELRK"
1967. 3514
/gene="SCO3845"
/note="synonym: SCH69.15"
1967. 3514
/gene="SCO3845"
/note="SCH69.15, possible protein phosphatase, len: 515aa;
similar to many of undefined function egs. TR:P71588
(EMBL:Z80233) hypothetical protein from Mycobacterium
tuberculosis (514 aa) fasta scores; opt: 989, z-score:
916.0, E(): 0. (42.9% identity in 536 aa overlap) and
TR:Q50188 (EMBL:Z70722) putative phosphoprotein phosphatase
from Mycobacterium leprae (509 aa) fasta scores; opt: 966,
z-score: 895.0, E(): 0. (40.0% identity in 530 aa
overlap). Contains 2 Pfam matches to entry PF00481 PP2C,
protein phosphatase 2C. could have a role in signalling."
/codon_start=1
/transl_table=11
/product="putative protein phosphatase"
/protein_id="CAB45212.1"
/db_xref="GI:5102797"
/db_xref="GOA:Q9XA19"
/db_xref="SPTREMBL:O9XA19"
/translacion="MSLSLRPAAGSHKGMIREGNSGYAGPRLAIADMGQQAAGE
VASSEVITVALDDVPGSDVLTSLGHAVORANDOLRQWVEEDPALEGMTTLTALL
WTQRLGVHVGDSRIVLRGLVLTQTDHTVORLVDEGRITEREAGTHPORSILIM
RALGSGDRVFDLSIREVRADGRYLICSDGSLGVYSHQTHMEDILASQGGQETVQELI
QALRGGVDNITIVADVLDTLGLSDTTPVVGAVANQMGDNGMLQTP
AGRAALGRPGGGEFGPGSDVTGFIAGDFDDYDFDVFPRKSRWLKRS
LYTALAALIGGGYGGWRWTQTYVTGTDNDLHLYRGIQDLAWYLSKQXKHDE
IELKLPYPYQOKLVEATIPEDGLNDRAKIELAVQASACKQARRTAETKNAKTE
EGEAGTGTGTTPASFTSKASPSPNPGSPAPESSESPSTTATPGSGPTLSBEQKV
VSLGCKO"
1979. 2368
/gene="SCO3845"
/note="Pfam match to entry PF00481 PP2C, Protein
phosphatase 2C, score 32.60, E-value 2.8e-08."
2474. 2656
/gene="SCO3845"
/note="Pfam match to entry PF00481 PP2C, Protein
phosphatase 2C, score 9.50, E-value 0.11."
3542. 4981
/gene="SCO3846"
/note="synonym: SCH69.16"
3542. 4981
/gene="SCO3846"
/note="SCH69.16, probable FtsW/RodA/Spove family cell
cycle protein, len: 479aa; similar to many egs.
SW:FRSW_MYCTU probable cell division protein FtsW from
Mycobacterium tuberculosis (469 aa) fasta scores; opt:
1367, z-score: 1568.6, E(): 0. (48.8% identity in 451 aa
overlap) and SW:SPSE_BACSU stage V sporulation protein E
from Bacillus subtilis (366 aa) fasta scores; opt: 496,
z-score: 572.9, E(): 1.4e-24, (31.7% identity in 391 aa
overlap). Also similar to SC6G9.31 from Streptomyces
coelicolor (446 aa) fasta scores; opt: 1132, z-score:
1182.0, E(): 0. (58.6% identity in 444 aa overlap).

```

```

Contains Pfam match to entry PF01098 FTSW RODA SPOVE, Cell
cycle protein. Also contains possible membrane spanning
hydrophobic regions."
/codon_start=1
/transl_table=11
/product="putative FtsW/RodA/Spove family cell cycle
protein"
/protein_id="CAB45213.1"
/db_xref="GI:5102798"
/db_xref="SPTREMBL:O9XA18"
/translacion="MSSTNFTSTHTSTIGAIPSRRTLALIVFAVLIPVFAYAN
VLGAINDVPAGLLSYGFLGGLAGVHVVRKFPAPYADPLILPLATLNGLGLVIAIW
RLQSDRLQSSKLFVEAAPROMLYSALGVAVLFLRHRVLRQVTVYSMAGALV
LLLLPLVGLGHDFNGAKIWKIPGLGTLOGBEPAKIVLAVFAGYLMVKDALALAS
VGLMLXAVAGVAVSFESHVQORVOAWLDPKHVELSRQGVFGHTESQMQLWAPGS
GGTLGSGWQGNSDILGFAANSDFILATFGBELGLAGLMLALLLIVLIRGVRITALA
ARDPFGKLLATLGSALFALQVVFAGVGMGLIPLTGMTMPLFAYGSGSVIANWALIGI
LLRISDTRRFPAPAPANPDAEMTQVRP"
3788. 4921
/gene="SCO3846"
/note="Pfam match to entry PF01098 FTSW RODA SPOVE, Cell
cycle protein, score 253.10, E-value 3.7e-72."
4978. 6450
/gene="SCO3847"
/note="synonym: SCH69.17"
4978. 6450
/gene="SCO3847"
/note="SCH69.17, probable secreted penicillin-binding
protein, len: 490 aa; similar to many e.g. TR:P71586
(EMBL:Z80233) putative penicillin-binding protein from
Mycobacterium tuberculosis (491 aa) fasta scores; opt:
112.1 z-score: 1262.1, E(): 0. (39.3% identity in 499 aa
overlap) and SW:P8PC_BACSU penicillin-binding protein 3
from Bacillus subtilis (668 aa) fasta scores; opt: 447,
z-score: 502.5, E(): 1.2e-20, (29.0% identity in 386 aa
overlap). Also similar to SC6G9.32 from Streptomyces
coelicolor (485 aa) fasta scores; opt: 1493, z-score:
1531.4, E(): 0. (49.4% identity in 488 aa overlap).
Contains Pfam match to entry PF00905 transpeptidase,
penicillin binding protein transpeptidase domain. Also
contains possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="putative secreted penicillin-binding protein"
/protein_id="CAB45214.1"
/db_xref="GI:5102799"
/db_xref="GOA:Q9XA17"
/db_xref="SPTREMBL:O9XA17"
/translacion="VNKPLRRIALPGLLVLTLLIRDNWLVQVQADELKHENRVA
IRYASPRGDIIVDGKATGHATTEGDFKXKTKYDGMWAPVTVGSVAFGATQLES
IDGILITDNDRLDFRNTLDMITGKRGNNVVTTLGAAQAAIDGLKQKAGAVV
AIEPSTGKILSMASYPYSDPTVIAGSNEAAGWNNKLOKNNPDPMRLALREVTPP
GSTFKVVTAAALEHGYASADEKTSPLPMTMPTGDTLKNVGNIPCNATILREALR
VSCNTVFGKLYDLGLNKMLETAKFGTEQFVFRSSASVFSMDMSQVALSSIG
QFNATATPLQAMVTSIAINNGTLMKPYMVDLOGLDITVEKTDPEENSEPLSADNA
QILQSMETVVEDGTGNAKIDGVKVGKTAQHGVDNSENYPYAFISYAKADDGSS
PYAVAVIEDENAVRDDISGGSLAAPIAKWMEAVIDSKK"
5395. 6417

```

```

misc_feature
Query Match 11.4%; Score 59.4; DB 1; Length 303550;
Best Local Similarity 46.7%; Pred. No. 3;
Matches 221; Conservative 0; Mismatches 251; Indels 1; Gaps 1;

Qy 48 TGCCTACGCGCGGTGACCGAGTGCACAGGTGGAGTTGACGAGCCAGCTCATCGCC 107
Db 289969 TGCCTACGCGCGGTGACCGAGTGCACAGGTGGAGTTGACGAGCCAGCTCATCGCC 289910
Qy 108 GCCTGCGCGCGGTGACCGAGTGCACAGGTGGAGTTGACGAGCCAGCTCATCGCC 167
Db 289909 GCCTGCGCGCGGTGACCGAGTGCACAGGTGGAGTTGACGAGCCAGCTCATCGCC 289850
Qy 168 TCGATCTGACGGTTCGCGCGCGCTCGATCTCGTTCGCGCGCGCTCCACACACCGAGGGAAC 227

```

```

Db 289849 GCAGCGCCCGCCCGCCAGCTCGGCTCTCTCCGCGCGAGGTCGTTGAAG 289790
Qy 228 TGCGTGGTGGAGCGGACCCAGCCAGCGGAGAGCTGCTCAGGCTGGGTGACGAGCC 287
Db 289789 GTACGCGCCAGACGCGCTCCCGCCCGCGCGTCCGTCGCCGAGCAGCGTCCCGGCATC 289730
Qy 288 TCTCGCGGCTCGGAGCGGAAACGCGACGCGCTCAGATCCCTGTCAGTGGCATCGGCTCA 347
Db 289729 CCGGCAGCGCCGCGCCGACGCGCAGCGTCCCGCGGCGCCGCGCGGACAGATCG 289670
Qy 348 GTGCGCGTCTGTCCTTGGCTTGGAGATAGCGTTTACACAGAGCGGACACGAGCG 406
Db 289669 GCCCGCATCCCGGACCGCGCATCCCGCGCGCGCGCTCCCGCGCGCCCATGCCCC 289610
Qy 407 GGGCGGGCGGGCGGGCGGTTTCAGCGCATCCGTCGATGACGCGGCTCGGGGTTCGGGG 466
Db 289609 GGGAGCGGCGCGGGTTCAGTCCGCTCCCGGGAGCGTCCCGCGCGCGGCGGACCGGGTG 289550
Qy 467 CGGTGCGGAGATCCGTACCGCGCGGACCGCTCGGCGGACGCGCGCGCGGATC 519
Db 289549 CGCACCGCGTCCACGAGGCGGTCCCGCACGCGCTCCACCGCGCGGTTCGGGGTC 289497

```

```

RESULT 9
AC136100 129506 bp DNA linear HTG 29-OCT-2002
LOCUS Rattus_norvegicus clone CH230-462A11, *** SEQUENCING IN PROGRESS
DEFINITION *** 57 unordered pieces.
AC136100
VERSION AC136100.1 GI:24417971
KEYWORDS HTG; HTGS PHASE1
SOURCE Rattus_norvegicus (Norway rat)
ORGANISM Rattus_norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

```

1 (bases 1 to 129506)
Mazny,D., Marle, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Guaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshuher, L., Louisedge, H., Lozada, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartine, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A.,
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,
Morrison, N., Nguyen, N., Norris, S., Nwalemele, O., Okunolu, G.,
Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,
Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,
Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,

```

```

Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,
Rose, M., Rose, R., Ruiz, S., Sanders, W., Savary, G., Scherer, S.,
Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,
Sisson, I., Stiller, C.D., Smajs, D., Sneed, A., Sodergren, E.,
Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,
Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,
Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D.,
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,
Wardren, H., Wei, X., White, F., Williams, G., Willison, R., Wleczek, R.,
Wooden, R., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,
Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,
Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R.,
Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: KCTT
Center clone name: CH230-462A11
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 75161 bases at least Q40
Consensus quality: 78106 bases at least Q30
Consensus quality: 80540 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1370: contig of 1370 bp in length
* 1371: gap of unknown length
* 1471: contig of 1175 bp in length
* 2646: gap of unknown length
* 2746: contig of 1590 bp in length
* 4336: gap of unknown length
* 4436: contig of 1352 bp in length
* 5787: gap of unknown length
* 5888: contig of 1381 bp in length
* 7269: gap of unknown length
* 7369: contig of 1375 bp in length
* 8744: gap of unknown length
* 8844: contig of 1287 bp in length
* 10130: gap of unknown length
* 10231: contig of 1101 bp in length
* 11332: gap of unknown length
* 12854: contig of 1322 bp in length
* 12854: gap of unknown length
* 14524: contig of 1670 bp in length
* 14624: gap of unknown length
* 15984: contig of 1260 bp in length
* 15984: gap of unknown length
* 17089: contig of 1105 bp in length
* 17189: gap of unknown length
* 18532: contig of 1344 bp in length
* 18533: gap of unknown length

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```
18633 20007: contig of 1375 bp in length
* 20008 20107: gap of unknown length
* 20108 20165: contig of 1258 bp in length
* 21366 21465: gap of unknown length
* 21466 22492: contig of 1027 bp in length
* 22493 23648: contig of 1056 bp in length
* 23649 23748: gap of unknown length
* 23749 25101: contig of 1353 bp in length
* 25102 25201: gap of unknown length
* 25202 26496: contig of 1295 bp in length
* 26497 28596: gap of unknown length
* 28597 28676: contig of 1980 bp in length
* 28677 30152: contig of 1476 bp in length
* 30153 30252: gap of unknown length
* 30253 31814: contig of 1562 bp in length
* 31815 31914: gap of unknown length
* 31915 3302: contig of 1388 bp in length
* 3303 33402: gap of unknown length
* 33403 34583: contig of 1181 bp in length
* 34584 34683: gap of unknown length
* 34684 36574: contig of 1891 bp in length
* 36575 36674: gap of unknown length
* 36676 38144: contig of 1470 bp in length
* 38145 38244: gap of unknown length
* 38245 40273: contig of 2029 bp in length
* 40274 40373: gap of unknown length
* 40374 41883: contig of 1510 bp in length
* 41884 41983: gap of unknown length
* 41984 44125: contig of 2142 bp in length
* 44126 44225: gap of unknown length
* 44226 45503: contig of 2278 bp in length
* 45504 48604: gap of unknown length
* 48605 48746: contig of 2143 bp in length
* 48747 48846: gap of unknown length
* 48847 51418: contig of 2572 bp in length
* 51419 51518: gap of unknown length
* 51519 54977: contig of 3459 bp in length
* 54978 55077: gap of unknown length
* 55078 57445: contig of 2268 bp in length
* 57446 57445: gap of unknown length
* 57446 59735: contig of 2290 bp in length
* 59736 61781: contig of 1946 bp in length
* 61782 61881: gap of unknown length
* 61882 63845: contig of 1964 bp in length
* 63846 63945: gap of unknown length
* 63946 65199: contig of 1254 bp in length
* 65200 65299: gap of unknown length
* 65300 66555: contig of 1356 bp in length
* 66556 66755: gap of unknown length
* 66756 69589: contig of 2834 bp in length
* 69590 69589: gap of unknown length
* 69591 72230: contig of 2541 bp in length
* 72231 72330: gap of unknown length
* 72331 74766: contig of 2346 bp in length
* 74767 74766: gap of unknown length
* 74777 76884: contig of 2108 bp in length
* 76885 76984: gap of unknown length
* 76985 79581: contig of 2597 bp in length
* 79582 79681: gap of unknown length
* 79683 83305: contig of 3624 bp in length
* 83306 83405: gap of unknown length
* 83406 85985: contig of 2480 bp in length
* 85986 85985: gap of unknown length
* 85987 89565: contig of 3580 bp in length
* 89566 89565: gap of unknown length
* 89567 93062: contig of 3397 bp in length
* 93063 93162: gap of unknown length
* 93163 96275: contig of 3113 bp in length
* 96276 96375: gap of unknown length
* 96376 99096: contig of 2721 bp in length

Query Match      11.1%; Score 57.6; DB 2; Length 129506;
Best Local Similarity 42.9%; Pred. No. 6.8;
Matches 213; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 24 TCGCCCGCGCGCCCATACAGAACTGCGACTCGGGGGGGTACACAGGTGACAGGTGG 83
   |||
Db 16460 TCGCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 16519
   |||

QY 84 GAGTTCAGCAGCAGCTCATCGCGCGCTCGCGCGGATCCGAAACACACCGGCGCAGATC 143
   |||
Db 16520 CCGCNCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16579
   |||

QY 144 ACGTCAGCAGCGCGATGCGCGCTCGATCTCGACGGTTCGGCGCGGAGCTCGATCTGTCG 203
   |||
Db 16580 GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16639
   |||

QY 204 CCGGCTCCACACACAGCGGGAACCTGGCTCGGTGCGAGCGGCGCGCGCGCGCGCGCG 263
   |||
Db 16640 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16699
   |||

QY 264 TCGTCAGGCTCGCGCTACGAGCCTTTCGCGGGTTCGGGAGCGGAAACGCGCGCTCGAG 323
   |||
Db 16700 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16759
   |||

QY 324 ATCCCTGTTCAGTCGCTCGGCTCAGTGGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCG 383
   |||
Db 16760 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16819
   |||

QY 384 TTCACGACGAGCGCGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 443
   |||
Db 16820 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16879
   |||

QY 444 GACGCGCGCTCGCGGGTTCGGGCGGTTCGGGAGATCGTACGCGCGCGCGCGCGCGCGCG 503
   |||
Db 16880 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16939
   |||

QY 504 CAGCGCGCGCGCGGATCC 520
   |||
Db 16940 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 16956

RESULT 10
SCO939119/c
LOCUS          299050 bp      DNA      linear      BCT 11-FEB-2003
DEFINITION    Streptomyces coelicolor A3(2) complete genome; segment 16/29.
ACCESSION     AL939119 AL160331 AL353816 AL356612 AL357432 AL357524 AL358692
               AL358818 AL359988 AL391406 AL391541 AL392149 AL392150 AL392178
               AL645882
VERSION        AL939119.1 GI:24427855
KEYWORDS
SOURCE
ORGANISM      Streptomyces coelicolor A3(2)
               Streptomyces coelicolor A3(2)
               Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
               Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS
Bentley,S.D., Chater,K.F., Cerdano-Tarraga,A.M., Challis,G.L.,
Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
Huang,C.H., Kieser,T., Larke,L., Murphy,L., Oliver,K., O'Neill,S.,
Rabinowitz,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
Parkhill,J. and Hopwood,D.A.
Complete genome sequence of the model actinomycete Streptomyces
```

coelcolor A3(2)  
 Nature 417. (6885), 141-147 (2002)  
 21996410  
 12000953  
 2. (bases 1 to 299050)  
 Bentley, S.D.  
 Direct Submission  
 Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces  
 sequencing team, Sanger Institute, Wellcome Trust Genome Campus,  
 Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk  
 On or before Oct 30, 2002 this sequence version replaced  
 gi:7242748, gi:7636006, gi:8248766, gi:8249971, gi:8388712,  
 gi:8439477, gi:8894718, gi:9857143, gi:9886716, gi:10129746,  
 gi:11595528.

FEATURES  
 source  
 1. .299050  
 /organism="Streptomyces coelicolor A3(2)"  
 /mol\_type="genomic DNA"  
 /strain="A3(2)"  
 /db\_xref="taxon:100226"  
 200..1727  
 /genes="rRNA"  
 /genes="rRNA"  
 /notes="rRNA, 16S ribosomal RNA. Coordinates based on  
 comparison with rRNA from Streptomyces coelicolor  
 (EMBL:AL109848)."  
 2004..5126  
 /genes="rRNA"  
 2004..5126  
 /genes="rRNA"  
 /notes="rRNA, 23S ribosomal RNA. Coordinates based on  
 comparison with rRNA from Streptomyces coelicolor  
 (EMBL:AL109848)."  
 5228..5348  
 /genes="rRNA"  
 5228..5348  
 /genes="rRNA"  
 /notes="rRNA, 5S ribosomal RNA. Coordinates based on  
 comparison with rRNA from Streptomyces coelicolor  
 (EMBL:AL109848)."  
 complement(5412..6080)  
 /genes="SCO4123"  
 /note="synonym: SCD72A.09c"  
 complement(5412..6080)  
 /genes="SCO4123"  
 /note="SCD72A.09c, possible two-component system response  
 regulator, len: 222aa; similar to many eg. TR:086632  
 (EMBL:AL031155) possible two-component regulator from  
 Streptomyces coelicolor (222 aa) fasta scores; opt: 973,  
 z-score: 1128.5, E(): 0, 69.1% identity in 220 aa overlap.  
 Contains Pfam match to entry PF00072 response\_reg,  
 Response regulator receiver domain, Pfam match to entry  
 PF00196 GerE, Bacterial regulatory proteins, luxR family  
 and Prosite match to PS00622 Bacterial regulatory  
 proteins, luxR family signature."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative two-component system response  
 regulator"  
 /protein\_id="CAB92376.1"  
 /db\_xref="GI:8052435"  
 /db\_xref="GOA:Q9KYB7"  
 /db\_xref="SPTRMBL:Q9KYB7"  
 /translation="MTVLADQPLQRFQFMFLLESQDDMTVGEANGAARMAA  
 ELHPDVLMDRMPGLDGEATRIIVAGGRSKVLITFTDLDEYASIGRAGAGGFL  
 VKDALPEELSGSRATVADVAPSLTRLLDAYVHLPAAAGTRDHTPGSPDRLSP  
 LTERETELFVWGQGWNTIEAERLHAESTVTKVHGRILAKTGRSDRVQAVILAYET  
 KLVG"  
 complement(5415..5609)  
 /genes="SCO4123"  
 /note="Pfam match to entry PF00196 GerE, Bacterial  
 regulatory proteins, luxR family, score 61.00, E-value

2.6e-14"  
 complement(5475..5558)  
 /genes="SCO4123"  
 /note="PS00622 Bacterial regulatory proteins, luxR family  
 signature"  
 complement(5736..5077)  
 /genes="SCO4123"  
 /note="Pfam match to entry PF00072 response\_reg, Response  
 regulator receiver domain, score 126.30, E-value 5.7e-34"  
 complement(6077..7447)  
 /genes="SCO4123"  
 /note="synonym: SCD72A.10c"  
 complement(6077..7447)  
 /genes="SCO4123"  
 /note="SCD72A.10c, possible two-component sensor kinase,  
 len: 456aa; similar to many eg. TR:086631 (EMBL:AL031155)  
 putative two-component sensor kinase from Streptomyces  
 coelicolor (429 aa) fasta scores; opt: 929, z-score:  
 995.1, E(): 0, 42.7% identity in 422 aa overlap. Contains  
 Pfam match to entry PF00512 signal, Histidine kinase and  
 possible membrane-spanning hydrophobic regions. Contains  
 possible hydrophobic membrane spanning regions"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative two-component sensor kinase"  
 /protein\_id="CAB92377.1"  
 /db\_xref="GI:8052436"  
 /db\_xref="GOA:Q9KYB6"  
 /db\_xref="SPTRMBL:Q9KYB6"  
 /translation="MEATVIFGYAPHLYGPTDARPARTPDHGMPEVKVPEDRA  
 AEPTEYQVRNERRIRAFDRHFLLDLTGFWLAALVDGDMRSVAHDGPGVPV  
 GLVLVLSALSVFVWRRSHPRGALLITPAALVSANSGAALQALLQLLHIALR  
 LPRLNLWAVAVAAATCVSVYRGSAGSNQQLGSLVAIGVAALIGITVTRRGY  
 TEALDRARLETERDQOVLAAARARIAREMHDIIGHNLVITGLDGGVAAAK  
 SPDAQALDAIGTTSQALGELRLDLVLRDDQDGTGPPELAPQALTDRLDLD  
 GVRAGLPVETTVKGEPLPEGRQLTVRVIQEALNTLKHAGPDATAHLELYSGSG  
 AATLVTDGTGRTAPLSPSPGSGRLPGMRERTSLGGLLEAGRPAEQGWRVHLLH  
 PEESPPQ"  
 complement(6095..6670)  
 /genes="SCO4124"  
 /note="Pfam match to entry PF00512 signal, Histidine  
 kinase, score -41.40, E-value 0.0087"  
 7473..8006  
 /genes="SCO4125"  
 /note="synonym: SCD72A.11"  
 7473..8006  
 /genes="SCO4125"  
 /note="SCD72A.11, possible acetyltransferase, len: 177 aa;  
 weakly similar to many eg. TR:AD52985 (EMBL:AF140221)  
 aminoglycoside 6'-N-acetyltransferase from Xanthomonas  
 maltophilia (153 aa) fasta scores; opt: 128, z-score:  
 164.9, E(): 0.1, 28.7% identity in 115 aa overlap.  
 Contains Pfam match to entry PF00583 Acetyltransf,  
 Acetyltransferase (GNAT) family."  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative acetyltransferase"  
 /protein\_id="CAB92378.1"  
 /db\_xref="GI:8052437"  
 /db\_xref="GOA:Q9KYB5"  
 /db\_xref="SPTRMBL:Q9KYB5"  
 /translation="MIRSHADEWPAVKELRLRALRDPVADLAYLTYEEAARPDY  
 WQBARAGAGCAQVAVGGLWGLTTLVVEAGTTDWAGFVRRQGHVGV  
 FVREWFGSLTKALEFAGLENGWANGVERMLIVHFGNTAQQAGAYKAGFAPSGRTV  
 PLEGGPDHLELYVLER"  
 7476..7934  
 /genes="SCO4125"  
 /note="Pfam match to entry PF00583 Acetyltransf,  
 Acetyltransferase (GNAT) family, score 34.50, E-value  
 2.4e-06"  
 complement(8054..9541)  
 /genes="SCO4126"  
 /note="synonym: SCD72A.12c"

misc\_feature  
 gene  
 CDS



```
misc_feature 87808..96100
/notes="assembly fragment:01700
fragment chain:2"
misc_feature 96201..98287
/notes="assembly fragment:00028
fragment chain:2"
misc_feature 98388..102420
/notes="assembly fragment:00012
fragment chain:2"
misc_feature 102521..1168071
/notes="assembly fragment:00976
fragment chain:3"
misc_feature 168172..179602
/notes="assembly fragment:01864
fragment chain:3"
misc_feature 179703..181850
/notes="assembly fragment:01446"

ORIGIN
Query Match 10.8; Score 56.2; DB 2; Length 181850;
Best Local Similarity 48.78; Pred No 10;
Matches 243; Conservative 0; Mismatches 248; Indels 8; Gaps 3;

Qy 14 CGACCGGTGTGCGCCCGCGCGCCATACGAACTGCATCTCGCGGGGTGACCAAGTC 73
Db 4306 CCACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4365
Qy 74 GACCACTGGAGTGTGAGCAGCAGCTCATCGCGCTGCGCGGCATGCCGAAACACCG 133
Db 4366 GCGCGGGGGGGGGCGCGCGCTCGGGGGGGGGGGCGCGCGCGCGCGCGCGCGCGCG 4423
Qy 134 GGCACGATCAGCTGACGACGCGCATGCGCGCTCGATCTGACGCTCGCGCGCGAGTC 193
Db 4424 GCCAGGGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4483
Qy 194 GATCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253
Db 4484 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4543
Qy 254 CCGCGGACAGTGTCTCAGGTGCGGTTCACGACCTCTTCGCGGGTTCGCGAGCGAAACGCG 313
Db 4544 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4597
Qy 314 CACGCTCAGATCCTGTCTAGTCGATCGGCTAGTCGCGGTGTCTCCCTTGGCTGGG 373
Db 4598 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4657
Qy 374 AGGATAGCGGTTCACGACGCGCACACGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
Db 4658 GCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4717
Qy 434 TCCGCTCGATGACGAGCGGTGCGGGGTGCGGGGTGCGGGGTGCGGGGTGCGGGGT 493
Db 4718 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4777
Qy 494 CCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 512
Db 4778 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4796

RESULT 12
AC079420
LOCUS AC079420 239130 bp DNA linear HTG 01-SEP-2000
DEFINITION Mus musculus chromosome 15 clone RP23-104L12, WORKING DRAFT
SEQUENCE, 61 unordered pieces.
ACCESSION AC079420
VERSION AC079420.1 GI:9958032
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 239130)
```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

DOE Joint Genome Institute.  
Sequencing of Mouse  
Unpublished  
2 (bases 1 to 239130)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Project Name: 0  
Center clone name: RPCI-23\_104L12  
-----  
Summary Statistics  
Consensus quality: 154414 bases at least Q40  
Consensus quality: 183510 bases at least Q30  
Consensus quality: 194836 bases at least Q20  
Estimated insert size: 222300; agarose-fp estimation  
Estimated insert size: 23130; sum-of-contigs estimation  
Quality coverage: 3.27 in Q20 bases; agarose-fp estimation  
Quality coverage: 3.12 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 61 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1030: contig of 1030 bp in length  
\* 1031 1130: gap of unknown length  
\* 1131 2520: contig of 1390 bp in length  
\* 2521 2620: gap of unknown length  
\* 2621 3633: contig of 1013 bp in length  
\* 3634 3734: gap of unknown length  
\* 3735 5381: contig of 1648 bp in length  
\* 5382 5482: gap of unknown length  
\* 5483 6792: contig of 1311 bp in length  
\* 6793 6892: gap of unknown length  
\* 6893 8478: contig of 1586 bp in length  
\* 8479 8578: gap of unknown length  
\* 8579 9634: contig of 1056 bp in length  
\* 9635 9734: gap of unknown length  
\* 9735 10791: contig of 1057 bp in length  
\* 10792 10891: gap of unknown length  
\* 10892 12258: contig of 1367 bp in length  
\* 12259 12359: gap of unknown length  
\* 12360 13771: contig of 1413 bp in length  
\* 13772 13871: gap of unknown length  
\* 13872 14998: contig of 1127 bp in length  
\* 14999 15098: gap of unknown length  
\* 15099 16261: contig of 1163 bp in length  
\* 16262 16361: gap of unknown length  
\* 16362 17426: contig of 1065 bp in length  
\* 17427 17527: gap of unknown length  
\* 17528 18541: contig of 1015 bp in length  
\* 18542 18641: gap of unknown length  
\* 18642 20200: contig of 1559 bp in length  
\* 20201 20300: gap of unknown length  
\* 20301 21423: contig of 1123 bp in length  
\* 21424 21523: gap of unknown length  
\* 21524 22553: contig of 1030 bp in length  
\* 22554 22853: gap of unknown length  
\* 22854 23664: contig of 1010 bp in length  
\* 23665 23764: gap of unknown length  
\* 23765 24815: contig of 1052 bp in length  
\* 24816 24916: gap of unknown length  
\* 24917 26356: contig of 1441 bp in length  
\* 26357 26456: gap of unknown length



*	26457:	27956:	contig	of 1500	bp in length
*	27957	28056:	gap	of unknown	length
*	28057	29865:	contig	of 1809	bp in length
*	29866	29965:	gap	of unknown	length
*	29966	31409:	contig	of 1444	bp in length
*	31410	31509:	gap	of unknown	length
*	31510	34198:	contig	of 2889	bp in length
*	34199	34298:	gap	of unknown	length
*	34299	36147:	contig	of 1849	bp in length
*	36148	36247:	gap	of unknown	length
*	36248	37098:	contig	of 1741	bp in length
*	37989	38098:	gap	of unknown	length
*	38099	39156:	contig	of 1068	bp in length
*	39157	39256:	gap	of unknown	length
*	39257	43172:	contig	of 3916	bp in length
*	43173	43272:	gap	of unknown	length
*	43273	44382:	contig	of 1110	bp in length
*	44383	44482:	gap	of unknown	length
*	44483	46836:	contig	of 2344	bp in length
*	46827	46926:	gap	of unknown	length
*	46927	49015:	contig	of 2089	bp in length
*	49016	49115:	gap	of unknown	length
*	49116	51007:	contig	of 1892	bp in length
*	51008	51107:	gap	of unknown	length
*	51109	53919:	contig	of 2812	bp in length
*	53920	54019:	gap	of unknown	length
*	54020	56108:	contig	of 2089	bp in length
*	56109	56208:	gap	of unknown	length
*	56209	59412:	contig	of 3204	bp in length
*	59413	59512:	gap	of unknown	length
*	59513	62565:	contig	of 3053	bp in length
*	62566	64626:	gap	of unknown	length
*	64627	66266:	contig	of 2059	bp in length
*	64725	64824:	gap	of unknown	length
*	64825	70728:	contig	of 5904	bp in length
*	70729	70828:	gap	of unknown	length
*	70829	74333:	contig	of 3505	bp in length
*	74334	74433:	gap	of unknown	length
*	74434	80095:	contig	of 5662	bp in length
*	80096	80195:	gap	of unknown	length
*	80196	84711:	contig	of 4518	bp in length
*	84712	84813:	gap	of unknown	length
*	84814	89091:	contig	of 4278	bp in length
*	89092	89191:	gap	of unknown	length
*	89192	94478:	contig	of 5287	bp in length
*	94479	94579:	gap	of unknown	length
*	94579	98798:	contig	of 4220	bp in length
*	98799	98898:	gap	of unknown	length
*	98899	103681:	contig	of 4783	bp in length
*	103682	103781:	gap	of unknown	length
*	103782	109617:	contig	of 5836	bp in length
*	109618	109717:	gap	of unknown	length
*	109718	115589:	contig	of 5872	bp in length
*	115590	115689:	gap	of unknown	length
*	115690	122207:	contig	of 6518	bp in length
*	122208	122307:	gap	of unknown	length
*	122308	129310:	contig	of 7003	bp in length
*	129311	129410:	gap	of unknown	length
*	129411	135067:	contig	of 5657	bp in length
*	135068	135167:	gap	of unknown	length
*	135168	142552:	contig	of 7395	bp in length
*	142553	142652:	gap	of unknown	length
*	142653	151374:	contig	of 8722	bp in length
*	151375	151474:	gap	of unknown	length
*	151475	159603:	contig	of 8129	bp in length
*	159604	159703:	gap	of unknown	length
*	159704	169275:	contig	of 9572	bp in length
*	169276	169375:	gap	of unknown	length
*	169376	176542:	contig	of 7167	bp in length
*	176543	176642:	gap	of unknown	length
*	176643	185118:	contig	of 8476	bp in length
*	185119	185218:	gap	of unknown	length
*	185219	195151:	contig	of 9933	bp in length



AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (04-DEC-2000) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
[E-mail: hattori@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170]  
COMMENT On Jun 16, 2003 this sequence version replaced gi:21623976.  
FEATURES  
    source  
        1..191589  
            /organism="Homo sapiens"  
            /mol\_type="genomic DNA"  
            /db\_xref="taxon:9606"  
            /chromosome="11"  
            /map="11q"  
            /clone="CND-2655K5"  
ORIGIN  
Query Match 10.8%; Score 56; DB 9; Length 191589;  
Best Local Similarity 48.6%; Pred. No. 11;  
Matches 245; Conservative 0; Mismatches 250; Indels 9; Gaps 3;  
QY 3 TGCAGCAGCAACGACCGGTGCTGCGCCCGCGCGCCATCACGAATGCCACTCGCGCGG 62  
Db ITCTGACCGGGGCGAGCGGAGCCCTCGCCGCTCGGCTGCTGAGCTCGGTAGC 49339  
QY 63 GTGACCAAGTTCGACAGTGGAGTTGAGCAGCGAGCTCATCGCGGCTGCGCGCATG 122  
Db GGCAGCGCGCGCGCGCGCGCGGTGGGCGAGCAGCGCGCGCGCGCGACCTCGAG 49379  
QY 123 CCGAAACACCGGGCGAGGATCATGTCAGCAGCGCGATGCGCGCTCGATTCGACGGTC 182  
Db TCCAGCAGCTGGCGG-----CCACAGCGCGCGCTCGCGCTCGCCCTCGCGCGCG 49224  
QY 183 GCGCGAGCTGATCTCGTCCCGCGGTCCACACAGCGGGAATGCTGCTGTCGAGC 242  
Db TCGCGCTGCGCTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 49164  
QY 243 GCGAGCGCGCGCGCGCGCGCTGTCAGCGTGGCGTCCAGCGACCTC-TCGCGGGTCGG 300  
Db CGGAGCG 49104  
QY 301 GAGCGGAACCGCGCGCGCTCAGATCCCTGTCACTCGCATCGGCTCAGTCCGGTCG 360  
Db GCGCGCGCGCTCG 49044  
QY 361 CCCTTGGCTGGAGGATAGCGTTACAGCAGCGCGCGCGCGCGCGCGCGCGCG 420  
Db TGGCGCGCGCGCGCGCGCGT--CAGGTGACCGCGCGCGCGCGCGCGCGCGCGCG 49043  
QY 421 CCGGTTCAGCGCATCCGCTCGATCAGCAGCGCGCTGCGGGGTGCGGGCGGTGCG 480  
Db GCGGTGCG 48985  
QY 481 CTTACCG 504  
Db CCG 48902

RESULT 14  
NSP561198 89153 bp DNA linear BCT 23-OCT-2003  
LOCUS Actinomadura sp. ATCC 39727 gene cluster for biosynthesis of  
DEFINITION glycopeptide antibiotic A40926, strain ATCC 39727.  
ACCESSION AJ561198  
VERSION AJ561198.1 GI:32487222  
KEYWORDS ABC transporter; acyltransferase; beta-hydroxylase; dalbavancin;  
dbv1; dbv10; dbv11; dbv12; dbv13; dbv14; dbv15; dbv16; dbv17;  
dbv18; dbv19; dbv2; dbv20; dbv21; dbv22; dbv23; dbv24; dbv25;  
dbv26; dbv27; dbv28; dbv29; dbv3; dbv30; dbv31; dbv32; dbv33;  
dbv34; dbv35; dbv36; dbv37; dbv5; dbv6; dbv7; dbv8; dbv9;  
DpgA protein; DpgB protein; DpgC protein; DpgD protein;

glycosyltransferase; halogenase; hexose oxidase; HpgT  
aminotransferase; hydroxymandelate oxidase; hydroxymandelate  
synthase; integral membrane ion antiporter; m ethyltransferase;  
mannosyltransferase; monooxygenase; non-ribosomal peptide  
synthetase; ORF1; ORF10; ORF11; ORF12; ORF13; ORF14; ORF15; ORF16;  
ORF17; ORF18; ORF2; ORF3; ORF4; ORF5; ORF6; ORF7; ORF8; ORF9; OxyA  
monooxygenase; OxyB monooxygenase; oxygenase OxyC; prephenate  
dehydrogenase; regulator LuxR family; regulator StrR family;  
response regulator; sensory kinase; type II thioesterase; Vany-type  
carboxypeptidase.  
Nonomuraea sp. ATCC 39727  
Nonomuraea sp. ATCC 39727  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptosporangineae; Streptosporangiaceae; Nonomuraea.  
1  
Sosio, M., Stinchi, S., Beltrametti, F., Lazzarini, A. and Donadio, S.  
The gene cluster for the biosynthesis of the glycopeptide  
antibiotic A40926 by nonomuraea species  
Chem. Biol. 10 (6), 541-549 (2003)  
22721464  
12837387  
2 (bases 1 to 89153)  
Donadio, S.  
Direct Submission  
Submitted (12-MAY-2003) Donadio S., Microbial Technologies, Vicuron  
Pharmaceuticals, via R. Lepetit 34, 21040 Gerenzano, ITALY  
Location/Qualifiers  
1..89153  
    /organism="Nonomuraea sp. ATCC 39727"  
    /mol\_type="genomic DNA"  
    /strain="ATCC 39727"  
    /db\_xref="taxon:93944"  
    /note="(Actinomadura) sp. ATCC 39727 = cf. Nonomuraea sp."  
    /complement(<1..141)  
    /function="putative Ser/Thr protein kinase"  
    /note="ORF1"  
    /outside dalbavancin cluster"  
    /codon\_start=1  
    /transl\_table=11  
    /product="hypothetical protein"  
    /protein\_id="CAD91191.1"  
    /db\_xref="GI:32487223"  
    /translation="MGSAELLPGDPARLGDYWLAGRLSGGGGQVVEAYDEGGRRVA  
IKI"  
    /complement(347..1207)  
    /function="putative hydrolase"  
    /note="ORF2"  
    /outside dalbavancin cluster"  
    /codon\_start=1  
    /transl\_table=11  
    /product="hypothetical protein"  
    /protein\_id="CAD91191.1"  
    /db\_xref="GI:32487224"  
    /translation="WTRHTTVQIQLPVYREAGDPANFTLVLLHGFSSAMFNLI  
ADLADSVHLIADPHVGFQCSAAPSVEPPYDFDKLTETLELDTNDRALYTHD  
CAPGLRIASBPVSAITQSGNVLDTGTFPEWVLFPAKDRATHEPARELLTA  
DATWQYTHGVADLLERVSQSWLDQAYLDRGNKEVQLQFWDQFNLGYPQFQ  
EYFTRRPPLVWAGANDEIFGAGAEAYARDLPDAEHLDDAGHFALETHGRETAGL  
IRDFLRRLP"  
    /complement(1797..2627)  
    /function="putative DNA binding protein"  
    /note="ORF3"  
    /outside dalbavancin cluster"  
    /codon\_start=1  
    /transl\_table=11  
    /product="hypothetical protein"  
    /protein\_id="CAD91192.1"  
    /db\_xref="GI:32487225"  
    /translation="MSRODELFLRSLRVRVAPDVGLPARAGRRVRGLRRELAL  
AAGVADYVTRMEQGRQSDQVLEAVAGVLRLTABQRHLRELIGRHHARPPPA  
PPRLRLMMLDMLGPALLMTGLVDILNMGALITDPEMRPERNTEIRWIFLD  
PPARDVFDNEVAVAAAWLRGALGRSGDDPRPQQLVDELSAKSDAPARWAGHRT  
YCTVGTGKLRHPEVGVMTLNWESFAPTADPDILVLYTAPITGSPSEKLAALAGCGVS





misc feature

Query Match:

```
Query Match 10.7%; Score 55.6; DB 1; Length 303550;
Best Local Similarity 48.4%; Pred. No. 11;
Matches 246; Conservative 0; Mismatches 254; Indels 8; Gaps 3;
```

8 CACGAACGACCGGTGGTCCGCCCGGCCCGCCATCACGAACGTGCCTCCGGCGGGGTGAC 67  
72105 CACGTCCGGCGGCACATAGCGCAGCGGCGCGCTCGGCCCGCGGGGACCCCGGCGCTTCC 72164

68 CAGGTCGACCAGGTGGGAGTTGAGCAGCCAGCTCATCGCGGCTTGGCCGGCATGCCGA 127

72165 CACCAGGTGAGGGACCCGTCGTACCCGGCCGACCGATGGCCACCGCCACCGGCCGAG 72224

128 ACACCGGGCCAGGATCA CGTGCAGC CCGCCGATGGCGGCTTCGATCTCGACGCTGGCGG 187

[illegible]

72289

RESULT 16

AE067802S1	AB067802S1	780 bp	DNA	linear	PRI 10-OCT-2001
LOCUS	Homo sapiens gene for RNA binding protein HQK, exon 1.				
DEFINITION	AE067802				
ACCESSION					

```

VERSION      AB067802.1  GI:15991315
KEYWORDS

```

SEGMENT	1 of 8
SOURCE	Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

REFERENCE  
AUTHORS  
Li, Z. Z., Kondo, T., Murata, T., Ebersole, T. A., Nishi, T., Ushio, Y.,

**TITLE** Expression of a KH RNA binding protein, Hqk, is altered in human Yamamura, K. and Abe, K.

JOURNAL  
Unpublished  
glioma

REFERENCE  
AUTHORS  
2 (bases 1 to 780)  
1: 7 7 and the Y

TITLE	Direct Submission	Submitted (or was once submitted)
MEMBERS	22,912, and 200,000	

**JOURNAL**  
Submitted (05-AUG-2001) Kuniya Abe, Institute of Molecular  
Embryology and Genetics, Kumamoto University; Kuhonji 4-24-1,

Kumamoto, Kumamoto 862, Japan (E-mail: kabe@gpo.kumamoto-u.ac.jp,  
Tel: 81-96-373-6597 (ex. 6597) , Fax: 81-96-373-6597)

FEATURES	Location/Qualifiers
source	1. 780

```
/organism="Homo sapiens"  
/seq_type="contig" data-bbox="860 79 875 244"/>
```

```

/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

```

/chromosome="6"
/map="6q25-q27"

```

```

      /tissue_type="brain"
      <552. .693
      exon

```

/gene="Hqk"  
/product="RNA binding protein HOK"

```

/number=1
/note="CDS is reported in Acc#:AB067808"
/number=1

```

**ORIGIN**

T 191027 /

Query Match 10.6%; Score 55.2; DB 9; Length 780;

Best Local Similarity 48.4%; Pred. No. 55;  
Matches 183; Conservative 0; Mismatches 193; Indels 2; Gaps 1;

6 AGGACGACGACCGGTGGCGCGCGCGCCATACGACATCGACTCCGCGCGCGGTC 65

0 100 200 300 400 500 600 700 800 900 1000

DD  
141 AUGCCCCGCGGCACAGGGCGCGCGCGCGCTGAGCGGGCTGACCAAGCCGAGGAGCGGCGGC 200

QY 66 ACCAGGTCGACCGAGTGGAGTTGAGCAGCAGCTCATCGCCGCTGCGCCGCGATGCGG 125  
Db 201 CCGGGTCT 260  
QY 126 ABAACCGGCGCGAGTACGTCAGCAGCGGCGATGCGGCGCTCTCTCTCTCTCTCTCTCT 185  
Db 261 AGACCCGCT 320  
QY 186 CGCAGCTCGATCTGTCGCGCCGCTCTCCACACAGCGGGAATGCTGCTCGTGTGCGACGCG 245  
Db 321 AGCGCGCGCT 378  
QY 246 AGCCCGACCGGAGCAGCT 305  
Db 379 AAGCGCGCT 438  
QY 306 GAAACGCGCAGCGCTCAGATCCCTGTCTGTCGATCGGCTCTCTCTCTCTCTCTCTCTCT 365  
Db 439 GAGCGGCGCGCT 498  
QY 366 GGCCTGGAGATAGCGG 383  
Db 499 CTCGCGCGCGCGCGCGG 516

RESULT 17  
BX842572/c standard; circular genomic DNA; PRO; 341957 BP.

XX BX842572; AL021428; AL021926; AL021927; AL021928; AL021929;  
AC BX842572; AL021428; AL021926; AL021927; AL021928; AL021929;  
AC AL021930; AL123456; Z74410; Z80233; Z80775; Z86089; Z92669; Z92770; Z96071;  
AC Z97050;  
XX BX842572.1  
XX 21-NOV-2003 (Rel. 77, Created)  
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)  
DE Mycobacterium tuberculosis H37Rv complete genome; segment 1/13  
XX complete genome.  
XX Mycobacterium tuberculosis H37Rv  
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;  
OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.  
XX [1]  
EN MEDLINE; 98295987.  
RX PUBMED; 9634230.  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,  
RA Bagham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,  
RA Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,  
RA Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,  
RA Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,  
RA Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrall B.G.,  
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete  
RL genome sequence";  
RL Nature 393:537-544(1998).  
XX [2]  
RX PUBMED; 12368430.  
RA Camus J.C., Pryor M.J., Medigue C., Cole S.T.;  
RA "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";  
RL Microbiology 148:2967-2973 (2002).  
XX [3]  
RN 1-341957  
RP Parkhill J.;  
RT Submitted (11-JUN-1998) to the EMBL/GenBank/DBSJ databases.  
RL Submitted on behalf of the Mycobacterium tuberculosis sequencing and

mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut  
Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:  
parkhill@sanger.ac.uk  
XX Notes:  
CC Details of M. tuberculosis sequencing at the Sanger Centre  
CC are available on the World Wide Web.  
CC (URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/)  
XX Key Location/Qualifiers  
FT 1. 341957  
FT /db\_xref="taxon:83332"  
FT /mol\_type="genomic DNA"  
FT /organism="Mycobacterium tuberculosis H37Rv"  
FT /strain="H37Rv"  
FT 1. 1524  
FT /evidence=EXPERIMENTAL  
FT /note="Rv0001. (MT0001, MT029.01, P49993), len: 507 aa.  
FT dnaA, chromosomal replication initiator protein (see  
FT citations below), equivalent to other Mycobacterial  
FT CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g.  
FT P46388|DnaA MYC from Mycobacterium leprae (502 aa);  
FT Q9L7L7|DnaA MYC from Mycobacterium paratuberculosis (509  
FT aa); P49990|DnaA MYC from Mycobacterium avium (508 aa);  
FT P49992|DnaA MYC from Mycobacterium smegmatis (504 aa);  
FT etc. Also highly similar to others except in N-terminus  
FT e.g. Q9ZHT5|DnaA STRCH CHROMOSOMAL REPLICATION INITIATOR  
FT PROTEIN from Streptomyces chrysomallus (624 aa);  
FT Q9ZH76|DnaA STRE from Streptomyces reticuli (643 aa);  
FT DnaA ECOLI|P03004|B3702 chromosomal replication initiator  
FT protein from Escherichia coli strain K12 (467 aa), FASTA  
FT scores: opt: 986, E(): 0, (43.2% identity in 389 aa  
FT overlap); etc. Contains P50017 ATP/GTP-binding site motif  
FT A (P-loop) and P50108 DnaA protein signature. BELONGS TO  
FT THE DnaA FAMILY. Note that the first base of this gene has  
FT been taken as base 1 of the Mycobacterium tuberculosis  
FT H37Rv genomic sequence."  
FT /transl\_table=1  
FT /function="PLAYS AN IMPORTANT ROLE IN THE INITIATION AND  
FT REGULATION OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN  
FT OF REPLICATION. IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA  
FT AT A 9 BP CONSENSUS (DnaA BOX): 5'-TTATC(C/A)C(A/A)-3'.  
FT DnaA BINDS TO ATP AND TO ACIDIC PHOSPHOLIPIDS. DnaA PROTEIN  
FT BINDS THE ORIGIN OF REPLICATION (ORIC), ATP AND ADP, AND  
FT EXHIBITS WEAK ATPase ACTIVITY."  
FT /gene="dnaA"  
FT /locus\_tag="Rv0001"  
FT /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DnaA"  
FT /protein\_id="CAA16238.1"  
FT /translation="MIDDPGSGFTVMNVAVSELNGPKVDGSSDNLASPLTPQOR  
FT ARLNLQPLTIVEGFAALLSPVSPFVQNEIRHRAPIIDLRRLGHQIQGVRIAPPA  
FT TDEADTTVPSENPATSPDITDNDIIDAARAGDNQHSWPSYFTFRPHNTDSATA  
FT GVTSLNRYTFDTFVIGASRFHAALIAEAPARAYNPLFIWGESGLGTHLHAAG  
FT NYAQLFFPMVKVYKSTFETFTNDLRDRKVKRSYRDVLLVDDVDFTEGKEG  
FT IQEPEFTHTLNANKOIVSSDRPPKQATLEDRLTREFWGLITDVPQPELETRIA  
FT ILRKAQVERLAVPDVLELIASITERNIRELEGALJVTAFASINKTIDKALAEIVL  
FT RLIDADANTMQISAATVMAATAEFTDTEVIELRPGKTRALLAQSRQIAMIYLCRELTDL  
FT LFKIGQAFGRDHTTVMYAKRLISEMAERREVFDRVXELTTRIKQRKR"  
FT 622..645  
FT /note="P500017 ATP/GTP-binding site motif A (P-loop)"  
FT 1384..1440  
FT /note="P501008 DnaA protein signature"  
FT 2052..3260  
FT /evidence=EXPERIMENTAL  
FT /note="Rv0002, (MTV029.02, MTCY10H4.0), len: 402 aa. dnaN,  
FT DNA polymerase III (beta chain) (EC 2.7.7.7) (see citations  
FT below), equivalent to other Mycobacterial DNA POLYMERASES  
FT III BETA CHAIN e.g. NP 301130.1|NC 002677 from  
FT Mycobacterium leprae (399 aa); Q9L7L6|DP3B\_MYCPA from  
FT Mycobacterium avium subsp. paratuberculosis (399 aa);  
FT P52851|DP3B\_MYC from Mycobacterium smegmatis (397 aa);





Query Match		10.6%; Score 55; DB 11; Length 1393;
Best Local Similarity		42.9%; Pred. No. 50;
Matches 219; Conservative		0; Mismatches 291; Indels 1; Gaps 1;
QY	7	GCACGAACAGACCGTGGTGCCTCCCGCCGCGCCATACAGAACTCCCACTCCGCGCGGGTGA 66
DB	245	GCNCGCGCCNCANCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 305
QY	67	CGAGTGCACAGTGGAGTTAGACAGCAGCTCATCGCCGCTCATCGCGCGCGCGCGCGCG 126
DB	306	GNNGNGGGGNCNCG 365
QY	127	AACACCGGCGCAGGATCACGTGACAGACGCGCATGCGCGCTCGATCTCGACGCTCGGCC 186
DB	366	GCCGCGNNGCG 425
QY	187	GCAGCTCGATCTCGTCCCGCGCTCCACACAGAGGGAAGTCTGCTGTCGAGCGGCA 246
DB	426	GCGNNGCG 485
QY	247	GCCCCAGCGCGGACAGCTCGTCCAGGCTGGCGTCAAGGACCTCTCGCGGCTCGGAGCGG 306
DB	486	CGCCCCCG 545
QY	307	AAACGCG-CACGGTCAAGTCCCTGTAGTTCGATCGCATCGGCTCACTGCGCGTCTCCCTT 365
DB	546	CGGAACCGCCNCG 605
QY	366	GGCTCGGAGATAGCGGTTACGACAGAGCGGACACAGCGCGCGCGCGCGCGCGCGCG 425
DB	606	CACGCCCG 665
QY	426	TCAGCGGATCGGTCGATGACACAGCGGCTCGCGGCTCGCGGCGCGGTCGCGAGATCGGTAC 485
DB	666	CCCG 725
QY	486	CGCCGAGACCGCTCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
DB	726	GCG 756
RESULT 19		
AP004323/c		
LOCUS		95209 bp DNA linear HTG 21-MAR-2002
DEFINITION		Oryza sativa (japonica cultivar-group) chromosome 6 clone
ACCESSION		AP004323
VERSION		AP004323.1 GI:16930108
KEYWORDS		HTG; HTGS PHASE2
SOURCE		Oryza sativa (japonica cultivar-group)
ORGANISM		Oryza sativa (japonica cultivar-group)
REFERENCE		1 Sasaki, T., Matsumoto, T. and Yamamoto, K. Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, BAC clone: OJ1118.C02 Published Only in Database (2001)
AUTHORS		2 (Bases 1 to 95209) Sasaki, T., Matsumoto, T. and Yamamoto, K. Direct Submission Submitted (14-NOV-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail: tsasaki@nias.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468)
TITLE		The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.
JOURNAL		NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This
COMMENT		
FEATURES		source
ORIGIN		1..95209
Query Match		10.6%; Score 55; DB 2; Length 95209;
Best Local Similarity		37.6%; Pred. No. 18;
Matches 193; Conservative		0; Mismatches 318; Indels 2; Gaps 1;
QY	1	GGTGACGACGACGACGCGTGGTGCCTCCCGCGCGCCATACGAACTGCCACTCCGCGG 60
DB	69966	GNNGCGNNGNNGCG 69907
QY	61	GGGTGACGACGACGACGCGTGGGAGTTGAGCAGCAGCTCATCGCGCTCGCGCGGCA 120
DB	69906	GCCNNGNNGCG 69847
QY	121	TGCGGAAACACCGCGCGCAGGATCACGTGACAGCAGCGGATGCGCGCTTCGATCTCGACGG 180
DB	69846	CNNNGNNGCG 69787
QY	181	TGCGCGCGGACGTCGATCTCGTCCCGCGCGCTCCACACAGCGGAACTGGCTCGGTGCA 240
DB	69786	GCGNNGNNGCG 69727
QY	241	GCGCGAGCGCGCGCGGACAGCTCGTCCAGGCTGGCGTCAAGGCTCTCGCGGCTCGG 300
DB	69726	GCGCGNNGNNGCG 69667
QY	301	GAGCGGAAACGCGGACGCGCTCAGATCCCTGTGTCAGTTCGATCGGTCAGTTCGCGGTC 360
DB	69666	CGCGCG--CCGNNNCG 69609
QY	361	CCCTTGGCTGGAGAGTAGCGGTTTACGACGAGCGGACACCGCGCGCGCGCGCGCGCG 420
DB	69608	CG 69549
QY	421	GCGGTTTCAGCGGATCCGCTCGATGACAGCGGCTCGCGGCTCGCGGCGGTCGCGGAGATC 480
DB	69548	CCCG 69489
QY	481	CGTACCGCGCGGACCGCTCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
DB	69488	CG 69456
RESULT 20		
PM12H3G/c		
LOCUS		716 bp DNA linear STS 29-MAY-2003
DEFINITION		Penicillium marneffe STS, clone pm12h3.g, sequence tagged site.
ACCESSION		AL684460
VERSION		AL684460.1 GI:19337528
KEYWORDS		STS.
SOURCE		Penicillium marneffe
ORGANISM		Penicillium marneffe
REFERENCE		1 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.
AUTHORS		Yuen, K.Y., Pascal, G., Wong, S.S., Glaser, P., Woo, P.C., Kunst, F., Cai, J.J., Cheung, E.Y., Medigue, C. and Danchin, A.
TITLE		Exploring the Penicillium marneffe genome
JOURNAL		Arch. Microbiol. 179 (5), 339-353 (2003)

```

MEDLINE 22595073
PUBMED 12640520
REFERENCE 2 (bases 1 to 716)
AUTHORS Danchin,A. and Pascal,G.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong
FEATURES
source
1. 716
/organism="Penicillium marneffei"
/mol_type="genomic DNA"
/db_xref="taxon:37727"
/clone="pm12h3.g"

ORIGIN
Query Match 10.5%; Score 54.8; DB 11; Length 716;
Best Local Similarity 47.0%; Pred. No. 64;
Matches 207; Conservative 0; Mismatches 227; Indels 6; Gaps 1;

Qy 77 CAGTGGAGTTGAGCAGCCAGCTCATCGCGCGCTGCGCGCGCATGCGCGGCGCATGCGCGGCGC 136
Db 638 CGGCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579
Qy 137 CAGGATCAGTCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 196
Db 578 GAGCGGGGGGCACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519
Qy 197 CTGCTGCGCGCGCGCTCCACACAGCGGGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 256
Db 518 CTCGCCCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
Qy 257 GCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 316
Db 458 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399
Qy 317 GCCTCAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 376
Db 398 GGC-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 345
Qy 377 ATAGCGGTTACAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 436
Db 344 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 285
Qy 437 GCTGATGACAGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 496
Db 284 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 225
Qy 497 CTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 516
Db 224 CCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 205

RESULT 21
AF288517 4283 bp DNA linear PRI 02-AUG-2001
LOCUS Homo sapiens mutant insulin receptor substrate 2 (IRS2) gene,
partial cds.
ACCESSION AF288517
VERSION AF288517.1 GI:15077067
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4283)
AUTHORS Heyne,B., Gehrisch,S. and Jaross,W.
TITLES Two missense mutations in insulin receptor substrate 2 (G879S and
G882A)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4283)
AUTHORS Heyne,B., Gehrisch,S. and Jaross,W.
TITLES Direct Submission
JOURNAL Submitted (19-JUL-2000) Institut fuer Klinische Chemie und

Laboratoriumsmedizin, Universitaetsklinikum der Technischen
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany
FEATURES
Location/Qualifiers
1. 4283
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q34"
/note="obtained from a patient with diabetes mellitus"
139..>4283
/gene="IRS2"
4139..>4150
/gene="IRS2"
/product="mutant insulin receptor substrate 2"
139..>4150
/gene="IRS2"
/note="similar to NM_003749"
/codon_start=1
/product="mutant insulin receptor substrate 2"
/protein_id="AAK81053.1"
/db_xref="GI:15077068"
/translation="MSPPRHGPFGPAGSDGPNLNNNNNNNNHNSVRKQGYLRKQKHGH
KRFFVLRGAGGDEATAGGSAPOPRLVEYEGKWRKAGAPRVIALDCCCLNIN
KXADAKHYLIATYTKDRYFAVAENEQEGWYRALTDLVSEGAAGADAPAAAPA
ASCSASLPGALGGSAGAGADSGLVAPATAAEVYQVNLKPKGLGSKNLTGVYR
LCLSAITGFKVKNCEQPSVTQLQNMNIRCGHSDSPFIEVGRSAVTGGELWQADD
SVVAGNTHETILEAMKALKEFEPFRPSKSSGSSATHPIVSPGARHHLNLDPPS
QTGLVRRBTLSLAATPPAAKSCSRVTSSEGGDGAAGAAAGAPRVAVAGSPVLS
GEPVAPLRSHTLSGGCGRGSKVALLPAGGALQHSRSMSPVAHSPPAATSPGLSS
SSGCGSGVPPPPGPHPLPHLPKHGPPQPSGSGSASASGSPSDFGFMLOBYGSSPG
DLRACSHRSNTPEIAETPPARDGGGGGFYGVYMDRPLSHGCRSVRRVSGDAAQD
LDGRKRTYSLTTPAROPVPOPSASLDEVTLMRATPFGSAGRLCPSPASSPKVA
YHPYEDYDIEHGSRSSSNLGDADYMPMTFGAALAGSGSGSCRDDYMPSPAS
VBAQKLIQPRAAAAAAVPSAGPAGPAPTSAGRTFTPGCGGKFAHSPAESPEDS
GPMWCKSGKJSMHADKLLPNVDYLNVSFSDAVTTGPTDFFAALHPGGEPLRGV
PCCYSILPSRYKAPTCCGSDSDYQLMSSPVRIIEERLEPQTGPOQSAFPA
GCTQTPHPVPSVPSGSGPFLAQGRVAPRTRLSLEGSLPSPMHEVPLPEPK
SPGVINDFGSPGKALGPAPPLASASSSLLSPEASSPVPLPPRPSASSSLQ
SPLSYNMLDSSPKSPKXGHPGVGSLDGLLSPEASSPVPLPPRPSASSSLQ
PPPPAPGELYRLPPASAVATAQFGAASSLSDTGNDGYTEVAFVATPPQPIA
APFPEAKRVATSGVXRLSLMEQVSGVEAFLOASQPPDHRGAKYTRADPQGGRR
HSEFTSFTTTFVSPFAHNPKHNSASVENLSRKSSEGGVGGVGGGDEPTSP
RQLQAPPLAPQGREWTCQGLGVCFGSGSPMRRTSAGFQNGLYNIADVREEP
GLPPOPPLPQPPDQKSGMGRTRSLGLISAVGVSGTGGCGGCGPGPGLPPANTY
ASIDFLSHLKEATIVK"
2773
/gene="IRS2"
/note="compared to IRS2 deposited in GenBank Accession
Number AF322114; mutant encodes serine instead of wildtype
glycine"
/replaces="g"
2783
/gene="IRS2"
/note="compared to IRS2 deposited in GenBank Accession
Number AF322114; mutant encodes alanine instead of
wildtype glycine"
/replaces="g"

variation
variation
variation

ORIGIN
Query Match 10.5%; Score 54.8; DB 9; Length 4283;
Best Local Similarity 46.2%; Pred. No. 41;
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;

Qy 46 ACTGCCACTCCGCGCGGGTGACCCAGGTGGAGTTGAGCAGCCAGTCTCATCG 105
Db 1034 AGTTCCGCGCGCGCAGTAAAGACCAATCTCGGGGTCTGTCGGCCACGCCATCAGCG 1093
Qy 106 CGCGCTGCCCGCATGCCGAACCGCGCCAGGATCAGTCAGCAGCGGATCGCG 165
Db 1094 TCCCGCGCGCGCGCGCCACACACACCTGCTCAACCTGCCCCCAGCAGAGCGGCTCG 1153
Qy 166 CCGCGATCTCGACGCTCGCGCGCGCAGCTCGATCTCTGTCGCGCGGCTCCACACCGAGGGAA 225
```



```

Db 1154 TGGCGCGCTCGGCGACCGACAGCTGCGCGCCACCCCGCGGCGCAAGTGCAGCTCGT 1213
Qy 226 ACTGGCTCGGTGGCGAGCGGAGCCCGACGCGGGAAGCTCGTCCAGGCTGGGCTCACGGA 285
Db 1214 GCGGGTGGCGACCGCGAGCGGCGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1273
Qy 286 CCTCTCGCGGTGCGGAGCGGAAACGCGACGCGCTCAGATCCCTGTCAGTCCGATCGGCT 345
Db 1274 GCGCAGGCGGTGTGCTGGTGGGAGC-CCCTGAGCCCGCGGCGGTGCGCGCGCGCC 1332
Qy 346 CAGTGGCGGTGCTCCCTTGGCTGGGAGATAGGAGTTCACGACGAGCGGCACACCGG 405
Db 1333 CTGAGCGCTCGCACACCTGAGCGGCGGTGCGCGCGCGCGCGGAGCAAGTGGCGCTG 1392
Qy 406 CGGCGCGGCGGCGGCGGTTCAGCGGATCGGTCGATGACGACGCGCTCGCGGCTCGG 465
Db 1393 CTGCGCGGCGGCGGCGGTGCAACACGCGGCTCCATGTCCTGCGCGGCGCACTCG 1452
Qy 466 GCGGTGCGGAGATCGGTACCGCGCGGACCGCTCGGCGGCGCGCGCGCGG 515
Db 1453 CGCGCGCGCGCACCGCGCGGTCCCTGTCGTCAGCAGCGGCGCACCG 1502

```

```

RESULT 22
AF322114S1
LOCUS Homo sapiens insulin receptor substrate 2 (IRS2) gene, exon 1.
DEFINITION AF322114
ACCESSION AF322114
VERSION AF322114.1 GI:12247738
KEYWORDS
SOURCE 1 of 2
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4283)
AUTHORS Heyne,B.
TITLE Insulin receptor substrate 2 gene sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4283)
AUTHORS Heyne,B., Gehrisch,S. and Jaross,W.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2000) Institut fuer Klinische Chemie und
Laboratoriumsmedizin, Universitaetsklinikum der Technischen
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany
FEATURES
Location/Qualifiers
source
1..4283
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q34"
/notes="diabetes mellitus patient"
<1..4150
/gene="IRS2"
/number=1

```

```

exon
Query Match 10.5%; Score 54.8; DB 9; Length 4283;
Best Local Similarity 46.28; Pred. No. 41;
Matches 217; Conservative 0; Mismatches 252; Indels 1; Gaps 1;

Qy 46 ACTGCCACTCGGCGGCTGACCAAGTTCGACCAAGTGGGAGTTGAGCAGCTCATCG 105
Db 1034 AGTTCCGCGCGCAGTAAGACCAATCGTCGGGTCGTCGGCCACGCCACCCATCAGCG 1093
Qy 106 CCGCTCGCGCGGCTCCGAAACACCGGCGCAGGATCAGTGCAGCAGCGGATCGCG 165
Db 1094 TCCCGCGCGCGCGCGCACCAACCACTGTGTCACTGCCCGCCAGCAGCGGCGCTGG 1153
Qy 166 CCTCGATCTCGAGGTTCGGCGGAGCTCGATCTCTGTCGCGCGGCTCCACACCGAGGGA 225

```

```

Db 1154 TGGCGCGCTCGGCGACCGACAGCTGCGCGCCACCCCGCGGCGCAAGTGCAGCTCGT 1213
Qy 226 ACTGGCTCGGTGGCGAGCGGAGCCCGACGCGGGAAGCTCGTCCAGGCTGGGCTCACGGA 285
Db 1214 GCGGGTGGCGACCGCGAGCGGCGACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1273
Qy 286 CCTCTCGCGGTGCGGAGCGGAAACGCGACGCGCTCAGATCCCTGTCAGTCCGATCGGCT 345
Db 1274 GCGCAGGCGGTGTGCTGGTGGGAGC-CCCTGAGCCCGCGGCGGTGCGCGCGCGCC 1332
Qy 346 CAGTGGCGGTGCTCCCTTGGCTGGGAGATAGGAGTTCACGACGAGCGGCACACCGG 405
Db 1333 CTGAGCGCTCGCACACCTGAGCGGCGGTGCGCGCGCGCGGAGCAAGTGGCGCTG 1392
Qy 406 CGGCGCGGCGGCGGCGGTTCAGCGGATCGGTCGATGACGACGCGCTCGCGGCTCGG 465
Db 1393 CTGCGCGGCGGCGGCGGTGCAACACGCGGCTCCATGTCCTGCGCGGCGCACTCG 1452
Qy 466 GCGGTGCGGAGATCGGTACCGCGCGGACCGCTCGGCGGCGCGCGCGCGG 515
Db 1453 CGCGCGCGCGCACCGCGCGGTCCCTGTCGTCAGCAGCGGCGCACCG 1502

```

```

RESULT 23
AF385933
LOCUS Homo sapiens insulin receptor substrate 2 insertion mutant (IRS2)
DEFINITION AF385933
ACCESSION AF385933
VERSION AF385933.1 GI:14537856
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4286)
AUTHORS Heyne,B.
TITLE Two insertions in insulin receptor substrate 2 (N28_H29insN;
N28_H29insNN)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4286)
AUTHORS Heyne,B., Gehrisch,S. and Jaross,W.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) Institut fuer Klinische Chemie und
Laboratoriumsmedizin, Universitaetsklinikum der Technischen
Universitaet Dresden, Fetscherstr. 74, Dresden 01307, Germany
FEATURES
Location/Qualifiers
source
1..4286
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="13"
/map="13q34"
/notes="isolated from a patient with type 2 diabetes mellitus"
<139..>4153
/gene="IRS2"
<139..>4153
/gene="IRS2"
/product="insulin receptor substrate 2 insertion mutant"
139..>4153
/gene="IRS2"
/notes="N28_H29insN"
/codon_start=1
/product="insulin receptor substrate 2 insertion mutant"
/protein_id="AAK66751.1"
/db_xref="GI:14537857"
/translation="MASPPRHGPPGASGDGPNLNNNNNNNNHNSVRKCYLRKQKHG
HKRPVLRGPGAGDEATAGGSAPOPPREYVESEKKRSKAGAPKRVIALDCLNI
NKRADAKHLYLALYTKDEYFAVAENEQEGHYRALTDLVSEGRAAGDAPPAAP
AASCSLPGALGAGAGAGADSIGLVAPATRAYREVMQNLKPKGLGSKNLGYI
RLCUSARTIGVKLNCQSPSVTLQLMNIRCGHSDSFFFEVSGHSAVSGELMMQAD
DSVVAQNTIHETILEAMKALFELFPRRSKSSQSSGSATHPIISVFGARRHHLLVNLPP

```

```

gene
mRNA
CDS

```